PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00, C12Q 1/68 — A1

(11) International Publication Number: WO 98/13485

(43) International Publication Date: 2 April 1998 (02.04.98)

(21) International Application Number: PCT/US97/17302

(22) International Filing Date: 26 September 1997 (26.09.97)

(30) Priority Data:
60/037,742 27 September 1996 (27.09.96) US

(71) Applicant (*for all designated States except US*): MAXYGEN, INC. [US/US]; 3410 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US). VAN ES, Helmuth, H., G. [NL/NL]; Bandholm 89, NL-2133 DJ Hoofddorp (NL).

(74) Agents: SMITH, William, M. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION

(57) Abstract

The invention provides methods of evolving nucleic acids for use in gene therapy by recursive sequence recombination. Many of the methods evolve vectors, both viral and nonviral, to have improved properties. For example, vectors are evolved to have improved properties of viral titer, infectivity, expression of a gene within a vector, tissue specificity, viral genome capacity, episomal retention, lack of immunogenicity of the vectors or an expression product thereof, site-specific integration, increased stability, or capacity to confer cellular resistance to microorganism infection. The invention further provides an isolated O^6-methylguanine-DNA methyltransferase (MGMT) enzyme.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION

The present application is a Continuation-In-Part application ("CIP") of U.S. patent application serial no. ("USSN") 08/721,824, filed September 27, 1996, which was converted to Provisional application serial no. 60/037,742, under 35 U.S.C. § 111(b) and 37 C.F.R. § 1.53(b)(2); and a CIP of USSN 08/722,660, filed September 27, 1996. Each of the aforementioned applications is explicitly incorporated herein by reference in their entirety and for all purposes.

FIELD OF THE INVENTION

The present invention applies the field of molecular genetics to the improvement of vectors and other nucleic acids for use in gene therapy. Improvement is achieved by recursive sequence recombination.

BACKGROUND AND DESCRIPTION OF RELATED ART

Gene therapy is the introduction of a nucleic acid into cells of a patient to express the nucleic acid for some therapeutic purpose. That is, the nucleic acid is itself used as a drug. For example, an appropriate gene can be delivered to a patient with a recessive inherited disease, such as cystic fibrosis, to correct the genetic defect and cure the disease state. In other applications, delivery of genes encoding a toxin (e.g, diphtheria toxin, ricin, *tk*) can be used to kill cancer cells, and other genes can be specifically tailored to kill infectious organisms. Other applications include incorporation of regulatory sequences near endogenous genes. These different applications are directed to many different target cells with many modes of delivery (e.g., *in vitro, ex vivo, in situ*, intravenous, and germline modification).

The power of gene therapy has led many large pharmaceutical manufacturers and several smaller biotechnology companies to devote substantial financial and technical resources to developing gene therapy as a viable therapeutic approach to treating human

diseases. Although simple in theory, gene therapy is not without technical difficulties. Development of any gene therapy requires identification of a cell type as a target, means for entry of DNA into those cells, means for expressing useful levels of gene product over an appropriate time period, and avoidance of host immune response to the gene therapy agents.

The requirements for any particular application vary greatly and profoundly influence the choice of vector to be developed and tested. Possible variables in different applications include the efficacy of gene transfer, the efficacy of gene expression, the duration of gene expression, the feasibility of repeat dosing, and the ability to target appropriate cells and avoid inappropriate cells. Confounding factors that may arise include the inability of virus or delivery vehicle to enter into or integrate into the chromosomes of particular cells, the shutdown of transcriptional promoters, the loss of input DNA, the destruction of treated cells, and the neutralization of input virus or gene product. All of these factors depend on the choice of viral vector or non-viral delivery system and on the ability of the host to respond to that virus or delivery system.

Most of the components currently available for constructing gene therapy vectors were not evolved or developed for gene therapy, and thus may have many undesirable features and may lack efficacy in the desired gene therapy application. For example, most eukaryotic viruses have evolved to optimize virulence and viral reproduction, and most non-viral DNA delivery systems were designed to be used for experimental transfection in laboratory conditions, not for administration to humans.

Solutions to the above difficulties and inefficiencies are needed before gene therapy becomes effective for routine treatment of significant numbers of patients with common diseases. The present invention fulfills this and other needs by providing *inter alia* methods for improving vectors and other nucleic acids used in gene therapy by recursive sequence recombination.

SUMMARY OF THE INVENTION

The invention provides methods of evolving nucleic acids for use in gene therapy by recursive sequence recombination. The methods entail recombining at least first and second forms of the segment differing from each other in at least two nucleotides, to produce a library of recombinant segments. At least one recombinant segment from the library is then screened for a property useful in gene therapy. At least one recombinant

segment identified by the screening is then recombined with a further form of the segment, the same or different from the first and second forms, to produce a further library of recombinant segments. The further library is then screened to identify at least one further recombinant segment from the further library for improvement in the property useful for gene therapy. Further cycles of recombination and screening are performed as necessary until the further recombinant segment confers a desired level of the property useful for gene therapy.

In one embodiment, the invention provides for a method of modifying a nucleic acid segment for use in gene therapy by recursive sequence recombination, comprising the following steps: (1) recombining at least a first and a second form of the segment differing in at least two positions, to produce a first set of recombinant segments; (2) screening at least one recombinant segment for a property useful in gene therapy; (3) recombining at least one recombinant segment generated by steps (1) and (2) with a variant form of the segment, the same as or different from the first or second forms, to produce a second set of recombinant segments; and, (4) screening at least one recombinant segment from the second recombination set for the property useful for gene therapy. In a further embodiment of this method, steps (1) to (4) are repeated until the recursively recombined segment confers the property useful for gene therapy. In additional embodiments of this method the nucleic acid segment can be a viral nucleic acid segment, the viral nucleic acid segment can comprise a viral vector, or at least one recombining step occurs *in vivo* or *in vitro*.

In one embodiment, the desired property to be acquired is improved viral titer. Here, the recombinant segments are screened as components of viruses by propagation of the viruses on cells for multiple generations and isolation of progeny viruses, the progeny viruses being enriched for viruses having recombinant segments conferring the property of improved titer.

In a second embodiment, the desired property is improved viral infectivity. Recombinant segments can be screened as components of viruses by determining the percentage of a population of cells infected by a virus.

In a third embodiment, the desired property is improved expression of a gene within the nucleic acid segment. The recombinant segments can be screened by detecting expression of the recombinant segments within cells.

In a fourth embodiment, the desired property is improved or altered drug resistance. The recombinant segments can be screened by exposing the cells to the drug and selecting surviving cells, the surviving cells being enriched for recombinant segments having the property of improved or altered drug resistance.

In a fifth embodiment, the desired property is improved or altered tissue specificity. The recombinant segments can be screened as components of viruses by contacting the viruses with a first population of cells for which the property of infectivity by the virus is desired and a second population of cells for which the property of infectivity by the virus is not desired, and isolating progeny virus from the first population of cells, the progeny viruses being enriched for recombinant segments conferring the property of infectivity for the first subpopulation of cells.

In a sixth embodiment, the desired property is improved packaging capacity of a viral capsid. The recombinant segments can be screened as components of viruses by propagating the viruses on cells and isolating progeny viruses containing the recombinant segments. The packaging capacity of the viral capsid containing the recombinant segments is increased between successive screening steps.

In a seventh embodiment, the desired property is episomal retention. The cells containing the recombinant segments can be screened by propagating the cells without selection for the recombinant segments and then propagating the cells with selection for the recombinant segments, the cells surviving selection being enriched for cells harboring recombinant segments with the property of improved episomal retention.

In an eighth embodiment, the desired property is reduced immunogenicity of the recombinant segments or an expression product thereof. The recombinant segments can be screened by introducing the recombinant segments into a mammal and recovering surviving recombinant segments after a period of time.

In a ninth embodiment, the desired property is site-specific integration. The recombinant segments can be screened by introducing them into cells and recovering a region of cellular DNA including the desired site of integration, the region being enriched for recombinant segments with the property of site-specific integration.

In a tenth embodiment, the desired property is increased stability. The recombinant segments can be screened as components of viruses by subjecting the viruses to

destabilizing conditions and recovering surviving viruses, these viruses being enriched for recombinant segments conferring the property.

In an eleventh embodiment, the property is capacity to confer cellular resistance to microorganism infection. Cells containing recombinant segments can be screened for capacity to survive infection by the microorganism.

In a twelfth embodiment, the methods evolve vectors for introduction into target cells in nonviral form. Recombinant segments can be selected by introducing the recombinant segments into a mammal, recovering cells from the mammal into which the segments are integrated and are expressed to produce the protein or antisense RNA, and recovering the recombinant segments from the cells.

In a thirteenth embodiment, the invention provides methods of improving adenoassociated viral proteins *rep* and *cap* for expression in a packaging cell line. Cells containing recombinant segments of these genes are infected with a recombinant AAV (rAAV) containing a marker gene flanked by terminal repeat sequences (ITRs) and a helper virus, such as an adenovirus. The yield of progeny rAAV and helper virus produced by different cells are determined and cells having a high relative yield of rAAV to helper virus are selected.

In a fourteenth embodiment, the nucleic acid segment comprises a coding sequence encoding a protein or antisense RNA, which can be expressed after integration of the segment into genomic DNA of mammalian cells.

In a fifteenth embodiment, the nucleic acid segment encodes a viral protein and the property is capacity of a cell line containing the nucleic acid segment to package viral DNA transfected into the cell line.

In a sixteenth embodiment, the nucleic acid segment encodes a DNA binding protein, the property that is enhanced is uptake by a recipient cell of a vector encoding the DNA binding protein.

In a seventeenth embodiment, the invention provides an isolated recombinant O^6-methylguanine-DNA methyltransferase (MGMT) enzyme, as illustrated in Figure 5, with the amino acid sequence of SEQ ID NO:2, encoded by the nucleic sequence of SEQ ID NO:1. The enzyme can have at least one amino acid segment present in a natural human MGMT coding sequence and absent in a natural nonhuman MGMT coding sequence, and has at least

one amino acid segment present in the natural nonhuman MGMT coding sequence and absent in the natural human MGMT coding sequence. The enzyme can be a natural nonhuman MGMT coding sequence from mouse, rat, rabbit or hamster. The enzyme can be an isolated O^6-methylguanine-DNA methyltransferase (MGMT) enzyme comprising a protein encoded by SEQ ID NO:1. In alternative embodiments, the invention provides an expression vector comprising the O^6-methylguanine-DNA methyltransferase (MGMT) enzyme as shown in Figure 5 (SEQ ID NO:1), a host cell comprising this expression vector, and a transgenic animal comprising this expression vector.

In another embodiment, the invention provides a method of evolving a drug transporter gene, comprising: (1) recombinating at least first and second forms of the gene differing from each other in at least two nucleotides, to produce a library of recombinant genes;(2) screening at least one recombinant gene from the library for conferring improved or altered drug resistance; (3) recombining, as necessary, at least one recombinant gene with a further form of the gene, the same or different from the first and second forms, to produce a further library of recombinant genes; (4) screening, as appropriate, at least one further recombinant gene from the further library for improved or altered drug resistance; (5) repeating (3) and (4), as necessary, until the further recombinant gene confers a desired level of improved or altered drug resistance. In this method, more than one round of screening can be performed between successive steps of recombining. The recombinant or further recombinant genes are screened by exposing cells to a drug and selecting surviving cells, the surviving cells being enriched for recombinant or further recombinant genes having the property of conferring improved or altered drug resistance. These methods also can include increasing the concentration of the drug between successive rounds of screening. The drug can be a chemotherapeutic drug. In these methods, the recombinant or further recombinant genes can be screened by detecting efflux from cells of a substrate for a drug transporter encoded by the drug transporter gene or by the recombinant or further recombinant genes and selecting the cells containing low intracellular amounts of said substrate. The recombinant or further recombinant genes can be screened by detecting influx into cells of a substrate for a drug transporter encoded by the drug transporter gene or by the recombinant or further recombinant genes and selecting the cells containing high intracellular amounts of said substrate. In the methods, the cells can be stem cells, kidney cells, heart cells, lung cells, liver

cells, gastrointestinal or central nervous system cells. These methods can be for use of the recombinant or further recombinant gene in gene therapy. The method can have at least one recombining step occurring *in vivo*. The method can have at least one recombining step *in vitro*.

In another embodiment, the invention provides for a phagemid-adenovirus capable of generating single stranded DNA greater than 10 kilobases comprising an adenovirus and a phage f1 replication origin.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Scheme for *in vitro* shuffling, "recursive sequence recombination," of genes.

Figure 2: Scheme for selecting DNA binding proteins conferring enhanced DNA uptake by recipient cells.

Figure 3: Oligonucleotides used to generate recombinant forms of MGMT using the recursive recombination methods of the invention.

Figure 4: Illustrates the natural diversity of five known mammalian alkyltransferases - human, rat, mouse, hamster, and rabbit. This diversity was used to generate sequence diversity in the improved human MGMT gene.

Figure 5: Illustrates the nucleotide sequence (SEQ ID NO:1) and the amino acid sequence (SEQ ID NO:2) of the improved human MGMT gene generated by the methods of the invention.

Figure 6: Illustrates the construction of an novel adenovirus-phagmid.

DEFINITIONS

The term "screening" describes what is, in general, a two-step process in which one first determines which cells do and do not express a screening marker and then physically separates the cells having the desired property. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic circumstances, allows cells expressing the marker to survive while other cells die (or vice versa). Screening markers include luciferase, beta-galactosidase, and green fluorescent protein. Selection markers include drug and toxin

resistance genes. Although spontaneous selection can and does occur in the course of natural evolution, in the present methods selection is performed by man.

The term "exogenous DNA segment" refers to a DNA segment which is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

The term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins.

The terms "percentage sequence identity," "sequence identity," "sequence similarity" or "structural similarity" are calculated or determined by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison. Optimal alignment of sequences for aligning a comparison window can be conducted by computerized implementations of algorithms GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI.

The term "naturally-occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as is typical for the species.

The terms "isolated," "purified," or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state.

A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and,

where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

A specific binding affinity between two molecules, for example, a ligand and a receptor, means a preferential binding of one molecule for another in a mixture of molecules. The binding of the molecules can be considered specific if the binding affinity is about 1 x 10^4 M^{-1} to about 1 x 10^6 M^{-1} or greater.

Improved drug resistance is understood to mean resistance to a higher concentration of the drug, irrespective of the underlying process (such as higher affinity for the drug or increased pump activity).

Altered drug resistance is understood to mean any alteration in the drug resistance profile of a cell. This includes improved drug resistance, a change in the spectrum of drugs to which the cell shows resistance, and decreased drug resistance.

A stem cell is understood to mean a cell of the hematopoietic system that has the following characteristics: (1) it has the inherent ability to differentiate into any type of cell of the blood cell system, and (2) it has the capacity to multiply itself without loosing any of its inherent characteristics.

DETAILED DESCRIPTION

I. General

The invention provides methods of evolving, *i.e.*, modifying, a eic acid for the acquisition of or an improvement in a property or characteristic useful in gene therapy. The substrates for this modification, or evolution, vary in different applications, as does the property sought to be acquired or improved. Examples of candidate substrates for acquisition of a property or improvement in a property include viral and non nonviral vectors used in gene therapy. The methods require at least two variant forms of a starting substrate. The variant forms of candidate substrates can show substantial sequence or secondary structural similarity with each other, but they should also differ in at least two positions. The initial diversity between forms can be the result of natural variation, *e.g.*, the different variant forms (homologs) are obtained from different individuals or strains of an organism (including geographic variants) or constitute related sequences from the same organism (*e.g.*, allelic variations). Alternatively, the initial diversity can be induced, *e.g.*, the second variant form

can be generated by error-prone transcription, such as an error-prone PCR or use of a polymerase which lacks proof-reading activity (see Liao (1990) *Gene* 88:107-111), of the first variant form, or, by replication of the first form in a mutator strain (mutator host cells are discussed in further detail below). The initial diversity between substrates is greatly augmented in subsequent steps of recursive sequence recombination.

The properties or characteristics that can be sought to be acquired or improved vary widely, and, of course depend on the choice of substrate. For example, for viral and nonviral vector sequences, improvement goals include higher titer, more stable expression, improved stability, higher specificity targeting, higher frequency integration, reduced immunogenicity of the vector sequence or an expression product thereof, and higher expression of gene products. For genomic DNA from a packaging cell line used to package a viral vector used in gene therapy, the goals of improvement include increasing the titer of viruses produced by the cell line.

Improvement in a property or acquisition of a property is achieved by recursive sequence recombination. Recursive sequence recombination can be achieved in many different formats and permutations of formats, as described in further detail below. These formats share some common principles. Recursive sequence recombination entails successive cycles of recombination to generate molecular diversity. That is, create a family of nucleic acid molecules showing some sequence identity to each other but differing in the presence of mutations. In any given cycle, recombination can occur *in vivo* or *in vitro*, intracellular or extracellular. Furthermore, diversity resulting from recombination can be augmented in any cycle by applying prior methods of mutagenesis (e.g., error-prone PCR or cassette mutagenesis) to either the substrates or products for recombination. In some instances, a new or improved property or characteristic can be achieved after only a single cycle of *in vivo* or *in vitro* recombination, as when using different, variant forms of the sequence, as homologs from different individuals or strains of an organism, or related sequences from the same organism, as allelic variations..

A recombination cycle is usually followed by at least one cycle of screening or selection for molecules having a desired property or characteristic. If a recombination cycle is performed *in vitro*, the products of recombination, *i.e.*, recombinant segments, are sometimes introduced into cells before the screening step. Recombinant segments can also be

linked to an appropriate vector or other regulatory sequences before screening. Alternatively, products of recombination generated *in vitro* are sometimes packaged as viruses before screening. If recombination is performed *in vivo*, recombination products can sometimes be screened in the cells in which recombination occurred. In other applications, recombinant segments are extracted from the cells, and optionally packaged as viruses, before screening.

The nature of screening or selection depends on what property or characteristic is to be acquired or the property or characteristic for which improvement is sought, and many examples are discussed below. It is not usually necessary or desirable to understand the molecular basis by which particular products of recombination (recombinant segments) have acquired new or improved properties or characteristics relative to the starting substrates. For example, a gene therapy vector can have many component sequences each having a different intended role (*e.g.*, coding sequence, regulatory sequences, targeting sequences, stability-conferring sequences, and integration sequences). Each of these component sequences can be varied and recombined simultaneously. Screening/selection can then be performed, for example, for recombinant segments that have increased stable expression in a target cell without the need to attribute such improvement to any of the individual component sequences of the vector.

Initial round(s) of screening are often performed in bacterial cells due to high transfection efficiencies and ease of culture. Later rounds can be performed in mammalian cells to optimize recombinant segments for use in an environment close to that of their intended use. Final rounds of screening can be performed in the precise cell type of intended use (e.g., a stem cell). In some instances, this stem cell can be obtained from the patient to be treated with a view, for example, to minimizing problems of immunogenicity in this patient. In some methods, use of a gene therapy vector in treatment can itself be used as a round of screening. That is, gene therapy vectors that are successively taken up, integrated and/or expressed by the intended target cells in one patient are recovered from those target cells and used to treat another patient. The gene therapy vectors that are recovered from the intended target cells in one patient are enriched for vectors that have evolved, *i.e.*, have been modified by recursive recombination, toward improved or new properties or characteristics for specific uptake, integration and/or expression.

The screening or selection step identifies a subpopulation of recombinant segments that have evolved toward acquisition of a new or improved desired property or

properties useful in gene therapy. Depending on the screen, the recombinant segments can be identified as components of cells, components of viruses or in free form. More than one round of screening or selection can be performed after each round of recombination.

At least one and usually a collection of recombinant segments surviving screening/selection are subject to a further round of recombination. These recombinant segments can be recombined with each other or with exogenous segments representing the original substrates or further variants thereof. Again, recombination can proceed *in vitro* or *in vivo*. If the previous screening step identifies desired recombinant segments as components of cells, the components can be subjected to further recombination *in vivo*, or can be subjected to further recombination *in vitro*, or can be isolated before performing a round of *in vitro* recombination. Conversely, if the previous screening step identifies desired recombinant segments in naked form or as components of viruses, these segments can be introduced into cells to perform a round of *in vivo* recombination. The second round of recombination, irrespective how performed, generates further recombinant segments which encompass additional diversity than is present in recombinant segments resulting from previous rounds.

The second round of recombination can be followed by a further round of screening/selection according to the principles discussed above for the first round. The stringency of screening/selection can be increased between rounds. Also, the nature of the screen and the property being screened for can vary between rounds if improvement in more than one property is desired or if acquiring more than one new property is desired. Additional rounds of recombination and screening can then be performed until the recombinant segments have sufficiently evolved to acquire the desired new or improved property or function.

II. Formats for Recursive Sequence Recombination

Exemplary formats and examples for using recursive sequence recombination, sometimes referred to as DNA shuffling, sexual PCR or molecular breeding, have been described by the present inventors and co-workers in copending application United States Serial No. (USSN) 08/621,859, attorney docket no. 16528A-014612, filed March 25, 1996; international application PCT/US95/02126, filed February 17, 1995, published as WO 95/22625; Stemmer (1995) *Science* 270:1510; Stemmer (1995) *Gene* 164:49-53; Stemmer (1995) *Bio/Technology* 13:549-553; Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-

10751; Stemmer (1994) *Nature* 370:389-391; Crameri (1996) *Nature Medicine* 2:1-3; Crameri (1996) *Nature Biotechnology* 14:315-319.

(1) *In Vitro* Formats

One embodiment for shuffling DNA sequences *in vitro* is illustrated in Fig. 1. The initial substrates for recombination are a pool of related sequences, *e.g.*, different, variant forms, as homologs from different individuals or strains of an organism, or related sequences from the same organism, as allelic variations. The X's in the Fig. 1, panel A, show where the sequences diverge. The sequences can be DNA or RNA and can be of various lengths depending on the size of the gene or DNA fragment to be recombined or reassembled. Preferably the sequences are from 50 base pairs (bp) to 50 kilobases (kb).

The pool of related substrates are converted into overlapping fragments, *e.g.*, from about 5 bp to 5 kb or more, as shown in Fig. 1, panel B. Often, for example, the size of the fragments is from about 10 bp to 1000 bp, and sometimes the size of the DNA fragments is from about 100 bp to 500 bp. The conversion can be effected by a number of different methods, such as DNase I or RNAse digestion, random shearing or partial restriction enzyme digestion. For discussions of protocols for the isolation, manipulation, enzymatic digestion, and the like of nucleic acids, see, for example, Sambrook, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) (Sambrook), and, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. Greene Publishing and Wiley-Interscience, New York (1987) (Ausubel). The concentration of nucleic acid fragments of a particular length and sequence is often less than 0.1 % or 1% by weight of the total nucleic acid. The number of different specific nucleic acid fragments in the mixture is usually at least about 100, 500 or 1000.

The mixed population of nucleic acid fragments are converted to at least partially single-stranded form using a variety of techniques, including, for example, heating, chemical denaturation, use of DNA binding proteins, and the like. Conversion can be effected by heating to about 80°C to 100°C, more preferably from 90°C to 96°C, to form single-stranded nucleic acid fragments and then reannealing. Conversion can also be effected by treatment with single-stranded DNA binding protein (see Wold (1997) *Annu. Rev. Biochem.* 66:61-92) or recA protein (see Kiianitsa (1997) *Proc. Natl. Acad. Sci. USA* 94:7837-7840). Single-stranded nucleic acid fragments having regions of sequence identity

with other single-stranded nucleic acid fragments can then be reannealed by cooling to 20°C to 75°C, and preferably from 40°C to 65°C. Renaturation can be accelerated by the addition of polyethylene glycol (PEG), other volume-excluding reagents or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt concentration is from 10 mM to 100 mM. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%. The fragments that reanneal can be from different substrates as shown in Fig. 1, panel C. The annealed nucleic acid fragments are incubated in the presence of a nucleic acid polymerase, such as Taq or Klenow, and dNTP's (*i.e.* dATP, dCTP, dGTP and dTTP). If regions of sequence identity are large, Taq polymerase can be used with an annealing temperature of between 45-65°C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30°C. The polymerase can be added to the random nucleic acid fragments prior to annealing, simultaneously with annealing or after annealing.

The process of denaturation, renaturation and incubation in the presence of polymerase of overlapping fragments to generate a collection of polynucleotides containing different permutations of fragments is sometimes referred to as shuffling of the nucleic acid *in vitro*. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from 2 to 100 times, more preferably the sequence is repeated from 10 to 40 times. The resulting nucleic acids are a family of double-stranded polynucleotides of from about 50 bp to about 100 kb, preferably from 500 bp to 50 kb, as shown in Fig. 1, panel D. The population represents variants of the starting substrates showing substantial sequence identity thereto but also diverging at several positions. The population has many more members than the starting substrates. The population of fragments resulting from shuffling is used to transform host cells, optionally after cloning into a vector.

In one embodiment utilizing *in vitro* shuffling, subsequences of recombination substrates can be generated by amplifying the full-length sequences under conditions which produce a substantial fraction, typically at least 20 percent or more, of incompletely extended amplification products. Another embodiment uses random primers to prime the entire template DNA to generate less than full length amplification products. The amplification products, including the incompletely extended amplification products are denatured and subjected to at least one additional cycle of reannealing and amplification. This variation, in which at least one cycle of reannealing and amplification provides a

substantial fraction of incompletely extended products, is termed "stuttering." In the subsequent amplification round, the partially extended (less than full length) products reanneal to and prime extension on different sequence-related template species. In another embodiment, the conversion of substrates to fragments can be effected by partial PCR amplification of substrates.

In another embodiment, a mixture of fragments is spiked with one or more oligonucleotides. The oligonucleotides can be designed to include precharacterized mutations of a wildtype sequence, or sites of natural variations between individuals or species. The oligonucleotides also include sufficient sequence or structural homology flanking such mutations or variations to allow annealing with the wildtype fragments. Annealing temperatures can be adjusted depending on the length of homology.

In a further embodiment, recombination occurs in at least one cycle by template switching, such as when a DNA fragment derived from one template primes on the homologous position of a related but different template. Template switching can be induced by addition of recA (see Kiianitsa (1997) *supra*), rad51 (see Namsaraev (1997) *Mol. Cell. Biol.* 17:5359-5368), rad55 (see Clever (1997) *EMBO J.* 16:2535-2544), rad57 (see Sung (1997) *Genes Dev.* 11:1111-1121) or other polymerases (e.g., viral polymerases, reverse transcriptase) to the amplification mixture. Template switching can also be increased by increasing the DNA template concentration.

Another embodiment utilizes at least one cycle of amplification, which can be conducted using a collection of overlapping single-stranded DNA fragments of related sequence, and different lengths. Fragments can be prepared using a single stranded DNA phage, such as M13 (see Wang (1997) *Biochemistry* 36:9486-9492). Each fragment can hybridize to and prime polynucleotide chain extension of a second fragment from the collection, thus forming sequence-recombined polynucleotides. In a further variation, ssDNA fragments of variable length can be generated from a single primer by Pfu, Taq, Vent, Deep Vent, UlTma DNA polymerase or other DNA polymerases on a first DNA template (see Cline (1996) *Nucleic Acids Res.* 24:3546-3551). The single stranded DNA fragments are used as primers for a second, Kunkel-type template, consisting of a uracil-containing circular ssDNA. This results in multiple substitutions of the first template into the second. *See* Levichkin (1995) *Mol. Biology* 29:572-577; Jung (1992) *Gene* 121:17-24.

Reintroduction of Genes Shuffled *in vitro* into Cells

In a further embodiment, whole cells and organisms can be improved by evolving a transgene within those cells and organisms by recursive cycles of *in vitro* shuffling. The transgene is subjected to the recursive recombination methods of the invention, and the shuffled sequence library is put back into the cell/organism for selection. While this method is useful if multiple copies of the modified transgene are reintegrated into a cell, in a preferred variation of this selection assay, only a single copy of the modified transgene is inserted into each cell. Another preferred variation of this selection assay involves reducing the transcriptional expression variability of the modified transgene that may result from differences in chromosomal location of integration sites. This requires a means for defined, site-specific integration of the modified transgene. These methods can also be used to evolve an episomal vector (which can replicate inside the cell) which can site-specifically integrate into a chromosome.

Use of retroviruses to shuttle the modified transgene back into the cell for selection has the advantage that they integrate as a single copy. However, this insertion is not site-specific, *i.e.*, the retrovirus inserts in a random location in the chromosome. Adenoviruses and ars-plasmids are also used to shuttle modified transgenes, however, they integrate as multiple copies. While wild type AAV integrates as a single copy in chromosome q19, commonly used modified versions of AAV do not. Homologous recombination is also used to insert a modified recombinant segment (transgene) into a chromosome, but this method can be inefficient and may result in the integration of two copies in the pair of chromosomes. To solve these problems, one embodiment of the invention utilizes site-specific integration systems to target the transgene to a specific, constant location in the genome. A preferred embodiment uses the Cre/LoxP or the related FLP/FRT site-specific integration system. The Cre/LoxP system uses a Cre recombinase enzyme to mediate site-specific insertion and excision of viral or phage vectors into a specific palindromic 34 base pair sequence called a "LoxP site." Lox P sites can be inserted to a mammalian genome of choice, to create, for example, a transgenic animal containing the Lox P site, by homologous recombination (see Rohlmann (1996) *Nature Biotech.* 14:1562-1565). If a genome is engineered to contain a LoxP site in a desired location, infection of such cells with vectors carrying a gene for the Cre recombinase results in the efficient, site-specific

integration of the transgene-containing vector into the LoxP site. This approach is reproducible from cycle to cycle and provides a single copy of the modified transgene (recombinant sequence) at a constant, defined location. Thus, a transgene of interest can be modified using the recursive sequence recombination methods of the invention *in vitro* and reinserted into the cell for *in vivo/in situ* selection for the new or improved property in the optimal way with minimal noise. This technique can also be used *in vivo*, as discussed below. See, example, Agah (1997) *J. Clin. Invest.* 100:169-179; Akagi (1997) *Nucleic Acids Res.* 25:1766-1773; Xiao (1997) *Nucleic Acids Res* 25:2985-2991; Jiang (1997) *Curr Biol* 7:321-R323, Rohlmann (1996) *Nature Biotech.* 14:1562-1565; Siegal (1996) *Genetics* 144: 715-726; Wild (1996) *Gene* 179:181-188. The evolution of *Cre* is discussed in further detail, below.

(2) *In Vivo* Formats

(a) Plasmid-Plasmid Recombination

The recursive recombination methods of the invention include plasmid-plasmid recombinations. In this and other embodiments, the initial substrates for recombination are a collection of polynucleotides comprising variant forms of nucleic acid of interest, such as a gene, a vector, a transcriptional regulatory sequence, or the like. The variant forms can have substantial sequence identity to each other; for example, sequence identity sufficient to allow homologous recombination between substrates (see Datta (1997) *Proc. Natl. Acad. Sci. USA* 94:9757-9762; Shimizu (1997) *J. Mol. Biol.* 266:297-305; Watt (1985) *Proc. Natl. Acad. Sci. USA* 82:4768-4772). The diversity between the polynucleotides can be natural (*e.g.*, allelic or species variants), induced (*e.g.*, *in vitro* generated, as by error-prone PCR, see Light (1995) *Bioorg. Med. Chem.* 3:955-967), or the result of *in vitro* recombination. Diversity can also result from resynthesizing genes encoding natural proteins with alternative and/or mixed codon usage. There should be at least sufficient diversity between substrates that recombination can generate more diverse products than there are starting materials. There must be at least two substrates differing in at least two positions. However, in another embodiment, a library of substrates of 10^3-10^8 members is employed. The degree of diversity depends on the length of the substrate being recombined and the extent of the functional change to be evolved. Diversity at between 0.1-50% of positions is typical.

The diverse initial substrates or recombinant segments modified by the methods of the invention can be incorporated into plasmids. In one embodiment, the plasmids are standard cloning vectors, *e.g.*, bacterial multicopy plasmids. However, in alternative embodiments, described below, the plasmids include mobilization functions. The initial substrates or recombinant segments can be incorporated into the same or different plasmids. Often at least two different types of plasmid having different types of selection marker are used to allow selection for cells containing at least two types of vector. Also, where different types of plasmid are employed, the different plasmids can come from two distinct incompatibility groups to allow stable co-existence of two different plasmids within the cell. Nevertheless, plasmids from the same incompatibility group can still co-exist within the same cell for sufficient time to allow homologous recombination to occur.

Plasmids containing diverse substrates are initially introduced into procaryotic or eukaryotic cells by any transfection methods, *e.g.*, chemical transformation, natural competence, electroporation, viral transduction or biolistics (see, for example, Sambrook for a detailed descriptions of introducing DNA into cells; Hapala (1997) *Crit. Rev. Biotechnol.* 17:105-122). Often, the plasmids are present at or near saturating concentration (with respect to maximum transfection capacity) to increase the probability of more than one plasmid entering the same cell. The plasmids containing the various substrates or recombinant segments can be transfected simultaneously or in multiple rounds. For example, in the latter approach cells can be transfected with a first aliquot of plasmid, transfectants selected and propagated, and then infected with a second aliquot of plasmid.

Having introduced the plasmids into cells, recombination between substrates to generate recombinant genes or other nucleic acid segments occurs within cells containing multiple different plasmids merely by propagating the plasmids in the cells. However, cells that receive only one plasmid are unable to participate in recombination and the potential contribution of substrates on such plasmids to evolution (sequence modification) is not fully exploited, although these plasmids may contribute to new sequence diversity if they are propagated in mutator cells (described below) or otherwise accumulate point mutations (*i.e.*, by ultraviolet radiation treatment). The rate of evolution, *i.e.*, modification of nucleic acid sequence by the methods of the invention, can be increased by allowing all substrates to participate in recombination. In one embodiment, this is achieved by subjecting transfected

cells to electroporation. The conditions for electroporation are the same as those conventionally used for introducing exogenous DNA into cells (e.g., 1,000-2,500 volts, 400 μF and a 1-2 mM gap). Under these conditions, plasmids are exchanged between cells allowing all substrates to participate in recombination. In addition the products of recombination can undergo further rounds of recombination with each other or with the original substrate.

In another embodiment, the rate of evolution, i.e., the rate of recursive sequence modification, can also be increased by use of conjugative transfer. To exploit conjugative transfer, substrates are cloned into plasmids having MOB genes, and *tra* genes are also provided in cis or in trans to the MOB genes. The effect of conjugative transfer is very similar to electroporation in that it allows plasmids to move between cells and allows recombination between any substrate, and the products of previous recombination to occur merely by propagating the culture. The details of how conjugative transfer is exploited in these vectors are discussed in more detail below (see also Cabezon (1997) *Mol. Gen. Genet.* 254:400-406.)

The rate of evolution can also be increased by fusing cells to induce exchange of plasmids or chromosomes. Fusion can be induced by chemical agents, such as PEG, or viruses or viral proteins, such as influenza virus hemagglutinin, HSV-1 gB and gD, or fusigenic liposomes (see Dzau (1996) *Proc. Natl. Acad. Sci. USA* 93:11421-11425).

The rate of evolution can also be increased by use of mutator host cells; *e.g.*, bacterial Mut L, S, D, T, H mutator cells, insect (Drosophila) and mouse mutator cells, and human cell lines with defective DNA repair mechanisms, such as those from *Ataxia telangiectasia* patients, see Morgan (1997) *Cancer Res.* 57:3386-3389; Greener (1997) *Mol. Biotechnol.* 7:189-195; Mason (1997) *Genetics* 146:1381-1397; Aronshtam (1996) *Nucleic Acids Res* 24:2498-2504; Seong (1995) *Int. J. Radiat. Oncol. Biol. Phys.*33:869-874; Wu (1994) *J. Bacteriol.* 176:5393-5400; Rewinski (1987) *Nucleic Acids Res.*15:8205-8215; Aizawa (1986) *Jpn. J. Cancer Res.* 77:327-329.

The time for which cells are propagated and recombination is allowed to occur, of course, varies with the cell type but is generally not critical, because even a small degree of recombination can substantially increase diversity relative to the starting materials. Cells bearing plasmids containing recombined genes are subject to screening or selection for a desired function. For example, if the substrate being evolved contains a drug resistance

gene, one selects for drug resistance. In the case of drug resistance genes which encode drug transporteres flow cytometry can be employed to enrich for cells exhibiting high levels of a mutant transporter phenotype by screening for drug efflux. This is done by employing fluorescent transporter substrates or fluorescent analogues of the drug substrate in question. Specifically substrates that are poor substrates for the wildtype transporter are used. Sorting those cells exhibiting low levels of fluorescence will result in enrichment of cells expressing a mutant gene encoding a transporter pumping the substrate used. Cells surviving screening or selection or cells enriched by flow cytometry can be subjected to one or more rounds of screening/selection followed by recombination or can be subjected directly to an additional round of recombination.

The next round of recombination can be achieved by several different formats independently of the previous round. For example, in one embodiment, a further round of recombination can be effected simply by resuming (repeating) the electroporation or conjugation-mediated intercellular transfer of plasmids described above. In another embodiment, a fresh substrate or substrates, the same or different from previous substrates, can be transfected into cells surviving selection/screening. The new substrates can be included in plasmid vectors bearing a different selective (selection) marker(s) and/or from a different incompatibility group than the original plasmids. Selection markers confer a selectable phenotype on transformed cells. For example, the marker may encode antibiotic resistance, particularly resistance to chloramphenicol, kanamycin, G418, bleomycin and hygromycin, to permit selection of those cells transformed with the desired DNA sequences, see for example, Blondelet-Rouault (1997) Gene 190:315-317. Because selectable marker genes conferring resistance to substrates like neomycin or hygromycin can only be utilized in tissue culture, chemoresistance genes are also used as selectable markers *in vitro* and *in vivo*. Various target cells are rendered resistant to anticancer drugs by transfer of chemoresistance genes encoding P-glycoprotein, multidrug resistance-associated protein-transporter, dihydrofolate reductase, glutathione -S-transferase, O 6-alkylguanine DNA alkyltransferase (Tano (1997) *J. Biol. Chem.* 272:13250-13254), or aldehyde reductase (Licht (1997) *Stem Cells* 15:104-111) and the like.

As a further embodiment, cells surviving selection/screening can be subdivided into two subpopulations, and plasmid DNA from one subpopulation transfected

into the other, where the substrates from the plasmids from the two subpopulations undergo a further round of recombination. In either of the latter two embodiments, the rate of evolution can be increased by employing any of the techniques described above, including DNA extraction, electroporation, conjugation or use of mutator cells. In a still further embodiment, DNA from cells surviving screening/selection can be extracted and subjected to *in vitro* DNA shuffling.

After the second round of recombination, a further round of screening/ selection can be performed. In one embodiment, the screening or selection is performed under conditions of increased stringency. If desired, further rounds of recombination and selection/screening can be performed using the same strategies as used in the second round. With successive rounds of recombination and selection/screening, the surviving recombined substrates evolve toward acquisition of a desired phenotype or characteristic. Typically, in this and other recursive recombination methods of the invention, the final product of recombination that has acquired the desired phenotype can differ from starting (initial) substrates at 0.1%-25% of positions. The methods of the invention can evolve/modify nucleic acid sequences at a rate orders of magnitude in excess (*e.g.*, by at least 10-fold, 100-fold, 1000-fold, or 10,000 fold) of the rate calculated for naturally acquired mutation (about 1 mutation per 10^{-9} positions per generation, see Anderson (1996) *Proc. Natl. Acad. Sci. USA* 93:906-907).

(b) Virus-Plasmid Recombination

The recursive recombination methods of the invention include virus-plasmid recombinations. The strategy used for plasmid-plasmid recombination can also be used for other embodiments of the invention, including virus-plasmid recombination or phage-plasmid recombination. However, some additional comments particular to the use of viruses are appropriate. The initial substrates for recombination are cloned into both plasmid and viral vectors. It is usually not critical which substrate(s) are inserted into the viral vector and which into the plasmid, although usually the viral vector should contain different substrate(s) from the plasmid. As before, the plasmid (and the virus) typically contains a selective marker. The plasmid and viral vectors can both be introduced into cells by transfection as described above. However, a more efficient procedure is to transfect the cells with plasmid, select transfectants and infect the transfectants with virus. Because the efficiency of infection

of many viruses approaches 100% for many cells, most cells transfected and infected by this route contain both a plasmid and virus bearing different substrates.

Homologous recombination occurs between plasmid and virus generating both recombined plasmids and recombined virus. For some viruses, such as filamentous phage, in which intracellular DNA exists in both double-stranded and single-stranded forms, both can participate in recombination. Provided that the virus is not one that rapidly kills cells, recombination can be augmented by use of electroporation or conjugation to transfer plasmids between cells. Recombination can also be augmented for some types of virus by allowing the progeny virus from one cell to reinfect other cells. For some types of virus, virus infected-cells show resistance to superinfection. However, such resistance can be overcome by infecting at high multiplicity and/or using mutant strains of the virus in which resistance to superinfection is reduced.

The result of infecting plasmid-containing cells with virus depends on the nature of the virus. Some viruses, such as filamentous phage, stably exist with a plasmid in the cell and also extrude progeny phage from the cell (see Russel (1997) *Gene* 192:23-32). Other viruses, such as lambda having a cosmid genome, stably exist in a cell like plasmids without producing progeny virions. Other viruses, such as the T-phage and lytic lambda, undergo recombination with the plasmid but ultimately kill the host cell and destroy plasmid DNA. For viruses that infect cells without killing the host, cells containing recombinant plasmids and virus can be screened/selected using the same approach as for plasmid-plasmid recombination. Progeny virus extruded by cells surviving selection/screening can also be collected and used as substrates in subsequent rounds of recombination. For viruses that kill their host cells, recombinant genes resulting from recombination reside only in the progeny virus. If the screening or selective assay requires expression of recombinant genes in a cell, the recombinant genes should be transferred from the progeny virus to another vector, *e.g.*, a plasmid vector, and retransfected into cells before selection/screening is performed.

For filamentous phage, the products of recombination are present in both cells surviving recombination and in phage extruded from these cells. The dual source of recombinant products provides some additional options relative to the plasmid-plasmid recombination. In one embodiment, DNA can be isolated from phage particles for use in a round of *in vitro* recombination. In an alternative embodiment, the progeny phage can be

used to transfect or infect cells surviving a previous round of screening/selection, or fresh cells transfected with fresh substrates for recombination.

(c) Virus-Virus Recombination

The recursive recombination methods of the invention also include virus-virus recombinations. The principles described for plasmid-plasmid and plasmid-viral recombination can be applied to virus-virus recombination with a few modifications. The initial substrates for recombination are cloned into a viral vector. In a preferred embodiment, the same vector is used for all substrates. Preferably, the virus is one that, naturally or as a result of mutation, does not kill cells. After insertion, some viral genomes can be packaged *in vitro*. The packaged viruses are used to infect cells at high multiplicity such that there is a high probability that a cell receives multiple viruses bearing different substrates.

After the initial round of infection, subsequent steps depend on the nature of infection, as discussed in the previous section. For example, if the viruses have phagemid genomes such as lambda cosmids or M13, F1 or Fd phagemids, the phagemids behave as plasmids within the cell and undergo recombination simply by propagating the cells. Recombination is particularly efficient between single-stranded forms of intracellular DNA. Recombination can be augmented by electroporation of cells. Following selection/screening, cosmids containing recombinant genes can be recovered from surviving cells (*e.g.*, by heat induction of a cos$^-$ lysogenic host cell), repackaged *in vitro*, and used to infect fresh cells at high multiplicity for a further round of recombination.

If the viruses are filamentous phage, recombination of replicating form DNA occurs by propagating the culture of infected cells. Selection/screening identifies colonies of cells containing viral vectors having recombinant genes with improved properties, together with phage extruded from such cells. Subsequent options are essentially the same as for plasmid-viral recombination.

(d) Chromosome-Plasmid Recombination

The recursive recombination methods of the invention also include chromosome-plasmid recombinations. This format can be used to evolve both the chromosomal and plasmid-borne substrates. The format is particularly useful in situations in which many chromosomal genes contribute to a phenotype or one does not know the exact location of the chromosomal gene(s) to be evolved. The initial substrates for recombination are cloned into a plasmid vector. If the chromosomal gene(s) to be evolved are known, the

substrates constitute a family of sequences showing a high degree of sequence identity but some divergence from the chromosomal gene. If the chromosomal genes to be evolved have not been located, the initial substrates usually constitute a library of DNA segments of which only a small number show sequence identity to the gene or gene(s) to be evolved. Divergence between plasmid-borne substrate and the chromosomal gene(s) can be induced by mutagenesis or by obtaining the plasmid-borne substrates from a different species than that of the cells bearing the chromosome, as discussed above.

The plasmids bearing substrates for recombination are transfected into cells having chromosomal gene(s) to be evolved/ modified to acquire a new or modified property. Evolution by recursive recombination can occur simply by propagating the culture. In another embodiment, the nucleic acid sequence modification can be accelerated by transferring plasmids between cells by conjugation or electroporation. In a further embodiment, evolution by recursive recombination can be further accelerated by use of mutator host cells or by seeding a culture of nonmutator host cells being evolved with mutator host cells and inducing intercellular transfer of plasmids by electroporation or conjugation. Preferably, mutator host cells used for seeding contain a negative selection marker to facilitate isolation of a pure culture of the nonmutator cells being evolved. Selection/screening identifies cells bearing chromosomes and/or plasmids that have evolved toward acquisition or modification of a desired property or function.

Subsequent rounds of recombination and selection/screening proceed in similar fashion to those described for plasmid-plasmid recombination. For example, further recombination can be effected by propagating cells surviving recombination in combination with electroporation or conjugative transfer of plasmids. Alternatively, plasmids bearing additional substrates for recombination can be introduced into the surviving cells. Preferably, such plasmids are from a different incompatibility group and bear a different selective marker than the original plasmids to allow selection for cells containing at least two different plasmids. As a further alternative, plasmid and/or chromosomal DNA can be isolated from a subpopulation of surviving cells and transfected into a second subpopulation. Chromosomal DNA can be cloned into a plasmid vector before transfection.

(e) Virus-Chromosome Recombination

The recursive recombination methods of the invention also include chromosome-virus recombinations. As in previously described embodiments, the virus is

usually one that does not kill the cells, and is often a phage or phagemid. The procedure is substantially the same as for plasmid-chromosome recombination. Substrates for recombination are cloned into the vector. Vectors including the substrates can then be transfected into cells or *in vitro* packaged and introduced into cells by infection. Viral genomes recombine with host chromosomes merely by propagating a culture. Evolution can be accelerated by allowing intercellular transfer of viral genomes by electroporation, or reinfection of cells by progeny virions. Screening/selection identifies cells having chromosomes and/or viral genomes that have evolved toward acquisition of a new or modified property or desired function.

There are several options for subsequent rounds of recombination. For example, viral genomes can be transferred between cells surviving selection/recombination by electroporation. Alternatively, viruses extruded from cells surviving selection/screening can be pooled and used to superinfect the cells at high multiplicity. Alternatively, fresh substrates for recombination can be introduced into the cells, either on plasmid or viral vectors.

III. Vectors Used in Gene Therapy

The invention provides for methods of modifying a vector by recursive recombination for use in gene therapy. Broadly speaking, a gene therapy vector is an exogenous polynucleotide which produces a medically useful phenotypic effect upon the mammalian cell(s) into which it is transferred. A vector may or may not have an origin of replication. For example, it is useful to include an origin of replication in a vector for propagation of the vector prior to administration to a patient. However, the origin of replication can often be removed before administration if the vector is designed to integrate into host chromosomal DNA or bind to host mRNA or DNA. Vectors used in gene therapy can be viral or nonviral and include but are not restricted to those described for AAV vectors in patent applications PCT/NL96/00472 filed November 29 1996, for retrovirus vectors in patent application PCT/NL96/00195 filed May 7 1996 (published as WO96/35798), for adenovirus vectors in patent application EP-A-95202213 filed August 15 1995 and for nonvital gene transfer PCT/NL96/00324 filed August 16 1996.

Viral vectors are usually introduced into a patient as components of a virus. Illustrative vectors incorporating nucleic acids to be modified by the recursive recombination

methods of the invention include, for example, adenovirus-based vectors (Cantwell (1996) *Blood* 88:4676-4683; Ohashi (1997) *Proc Natl Acad Sci USA* 94:1287-1292), Epstein-Barr virus-based vectors (Mazda (1997) *J Immunol Methods* 204:143-151), adenovirus-associated virus vectors, Sindbis virus vectors (Strong (1997) *Gene Ther* 4: 624-627), herpes simplex virus vectors (Kennedy (1997) *Brain* 120: 1245-1259) and retroviral vectors (Schubert (1997) *Curr Eye Res* 16:656-662).

Nonviral vectors, typically dsDNA, can be transferred as naked DNA or associated with a transfer-enhancing vehicle, such as a receptor-recognition protein, liposome, lipoamine, or cationic lipid. This DNA can be transferred into a cell using a variety of techniques well known in the art. For example, naked DNA can be delivered by the use of liposomes which fuse with the cellular membrane or are endocytosed, *i.e.*, by employing ligands attached to the liposome, or attached directly to the DNA, that bind to surface membrane protein receptors of the cell resulting in endocytosis. Alternatively, the cells may be permeabilized to enhance transport of the DNA into the cell, without injuring the host cells. One can use a DNA binding protein, *e.g.*, HBGF-1, known to transport DNA into a cell. These procedures for delivering naked DNA to cells are useful *in vivo*. For example, by using liposomes, particularly where the liposome surface carries ligands specific for target cells, or are otherwise preferentially directed to a specific organ, one may provide for the introduction of the DNA into the target cells/organs *in vivo*.

A. Viral-Based Methods

Various viral vectors, such as retroviruses, adenoviruses, adenoassociated viruses and herpes viruses, are used in gene therapy. They are often made up of two components, a modified viral genome and a coat structure surrounding it (*see generally* Smith (1995) *Annu. Rev. Microbiol.* 49, 807-838), although sometimes viral vectors are introduced in naked form or coated with proteins other than viral proteins. Most current vectors have coat structures similar to a wildtype virus. This structure packages and protects the viral nucleic acid and provides the means to bind and enter target cells. However, the viral nucleic acid in a vector designed for gene therapy can be changed in many ways. The goals of these changes are to disable growth of the virus in target cells while maintaining its ability to grow in vector form in available packaging or helper cells, to provide space within the viral genome for insertion of exogenous DNA sequences, and to incorporate new sequences that

encode and enable appropriate expression of the gene of interest. Thus, vector nucleic acids generally comprise two components: essential *cis*-acting viral sequences for replication and packaging in a helper line and the transcription unit for the exogenous gene. Other viral functions are expressed in *trans* in a specific packaging or helper cell line.

(1) Retroviruses

Retroviruses comprise a large class of enveloped viruses that contain single--stranded RNA as the viral genome. During the normal viral life cycle, viral RNA is reverse-transcribed to yield double-stranded DNA that integrates into the host genome and is expressed over extended periods. As a result, infected cells shed virus continuously without apparent harm to the host cell. The viral genome is small (approximately 10 kb), and its prototypical organization is extremely simple, comprising three genes encoding gag, the group specific antigens or core proteins; pol, the reverse transcriptase; and env, the viral envelope protein. The termini of the RNA genome are called long terminal repeats (LTRs) and include promoter and enhancer activities and sequences involved in integration. The genome also includes a sequence required for packaging viral RNA and splice acceptor and donor sites for generation of the separate envelope mRNA. Most retroviruses can integrate only into replicating cells, although human immunodeficiency virus (HIV) appears to be an exception. This property can restrict the use of retroviruses as vectors for gene therapy.

Retrovirus vectors are relatively simple, containing the 5' and 3' LTRs, a packaging sequence, and a transcription unit composed of the gene or genes of interest, which is typically an expression cassette. Useful vectors have been described in PCT/NL96/00195 filed May 7 1996 disclosing vectors having mutant LTRs with the wildtype enhancer sequences replaced by a mutant polyoma enhancer sequence.

To grow such a vector, one must provide the missing viral functions in *trans* using a so-called packaging cell line. Such a cell is engineered to contain integrated copies of gag, pol, and env but to lack a packaging signal so that no helper virus sequences become encapsidated. Additional features added to or removed from the vector and packaging cell line reflect attempts to render the vectors more efficacious or reduce the possibility of contamination by helper virus.

The main advantage of retroviral vectors is that they integrate in the chromosome and are therefore potentially capable of long-term expression. They can be grown in relatively large amounts, but care is needed to ensure the absence of helper virus.

(2) Adenoviruses

Adenoviruses comprise a large class of nonenveloped viruses containing linear double-stranded DNA. The normal life cycle of the virus does not require dividing cells and involves productive infection in permissive cells during which large amounts of virus accumulate. The productive infection cycle takes about 32-36 hours in cell culture and comprises two phases, the early phase, prior to viral DNA synthesis, and the late phase, during which structural proteins and viral DNA are synthesized and assembled into virions. In general, adenovirus infections are associated with mild disease in humans.

Adenovirus vectors are somewhat larger and more complex than retrovirus or AAV vectors, partly because only a small fraction of the viral genome is removed from most current vectors. If additional genes are removed, they are provided in trans to produce the vector, which so far has proved difficult. Instead, two general types of adenovirus-based vectors have been studied, E3-deletion and E1-deletion vectors. Some viruses in laboratory stocks of wild-type lack the E3 region and can grow in the absence of helper. This ability does not mean that the E3 gene products are not necessary in the wild, only that replication in cultured cells does not require them. Deletion of the E3 region allows insertion of exogenous DNA sequences to yield vectors capable of productive infection and the transient synthesis of relatively large amounts of encoded protein.

Deletion of the E1 region disables the adenovirus, but such vectors can still be grown because there are several human cell lines (called 293, 911 and PER.C6) are available that constitutively express the E1 region of Ad5. Most recent gene therapy applications involving adenovirus have utilised E1 replacement vectors grown in PERC6 cells disclosed in PCT/NL96/00244 filed June 14 1996 (published as W097/00326). PerC6 produced recombinant adenovirus lots carrying for example the HSV thymidine kinase gene do not contain any detectable levels of replication competent adenovirus (RCA) and are therefore preferred for use in gene therapy and thus are an embodiment of the present invention.

The main advantages of adenovirus vectors are that they are capable of efficient episomal gene transfer in a wide range of cells and tissues and that they are easy to grow in large amounts. The main disadvantage is that the host response to the virus appears to limit the duration of expression and the ability to repeat dosing, at least with high doses of first-generation vectors.

In another embodiment, the recursive recombination methods of the invention are used to construct a novel adenovirus-phagmid capable of packaging DNA inserts over 10 kilobases in size. Incorporation of a phage f1 origin in a plasmid using the methods of the invention also generates a novel *in vivo* shuffling format capable of evolving whole genomes of viruses, such as the 36 kb family of human adenoviruses. The widely used human adenovirus type 5 (Ad5) has a genome size of 36 kb. It is difficult to shuffle this large genome *in vitro* without creating an excessive number of changes which may cause a high percentage of nonviable recombinant variants. To minimize this problem and achieve whole genome shuffling of Ad5, an adenovirus-phagemid was constructed. The invention's Ad-phagemid has been demonstrated to accept inserts as large as 15 and 24 kilobases and to effectively generate ssDNA of that size. In a further embodiment, larger DNA inserts, as large as 50 to 100 kb are inserted into the Ad-phagemid of the invention; with generation of full length ssDNA corresponding to those large inserts. Generation of such large ssDNA fragments provides a means to evolve, *i.e.* modify by the recursive recombination methods of the invention, entire viral genomes. Thus, this invention provides for the first time a unique phagemid system capable of cloning large DNA inserts (>10 KB) and generating ssDNA *in vitro* and *in vivo* corresponding to those large inserts.

The genomes of related serotypes of human adenovirus are shuffled *in vivo* using this unique phagmid system, as described in Example 4 and illustrated in Figure 6. The genomic DNA is first cloned into a phagemid vector, and the resulting plasmid, designated a "Admid," can be used to produce single-stranded (ss) Admid phage by using a helper M13 phage. To achieve *in vivo* recombination, ssAdmid phages containing the genome of homologous human adenoviruses are used to perform high multiplicity of infection (MOI) on F+ *mutS E. coli* cells. The ssDNA is a better substrate for recombination enzymes such as RecA. The high MOI ensures that the probability of having multiple cross-overs between copies of the infecting ssAdmid DNA is high. The shuffled adenovirus genome is generated by purification of the double stranded Admid DNA from the infected cells and is introduction into a permissive human cell line to produce the adenovirus library. This genomic shuffling strategy is useful for creation of recombinant adenovirus variants with changes in multiple genes. This allows screening or selection of recombinant variant phenotypes resulting from combinations of variations in multiple genes.

(3) Adeno-Associated Virus (AAV)

AAV is a small, simple, nonautonomous virus containing linear single-stranded DNA. See Muzycka, *Current Topics Microbiol. Immunol.* 158, 97-129 (1992). The virus requires co-infection with adenovirus or certain other viruses in order to replicate. AAV is widespread in the human population, as evidenced by antibodies to the virus, but it is not associated with any known disease. AAV genome organization is straightforward, comprising only two genes: *rep* and *cap*. The termini of the genome comprises terminal repeats (ITR) sequences of about 145 nucleotides.

AAV-based vectors typically contain only the ITR sequences flanking the transcription unit of interest. The length of the vector DNA cannot greatly exceed the viral genome length of 4680 nucleotides. Currently, growth of AAV vectors is cumbersome and involves introducing into the host cell not only the vector itself but also a plasmid encoding *rep* and *cap* to provide helper functions. The helper plasmid lacks ITRs and consequently cannot replicate and package. In addition, helper virus such as adenovirus is often required. The potential advantage of AAV vectors is that they appear capable of long-term expression in nondividing cells, possibly, though not necessarily, because the viral DNA integrates. The vectors are structurally simple, and they may therefore provoke less of a host-cell response than adenovirus. A major limitation at present is that AAV vectors are extremely difficult to grow in large amounts.

B. Non-Viral Gene Transfer Methods

Nonviral nucleic acid vectors used in gene therapy include plasmids, RNAs, antisense oligonucleotides (e.g., methylphosphonate or phosphorothiolate), polyamide nucleic acids, and yeast artificial chromosomes (YACs). Such vectors typically include an expression cassette for expressing a protein or RNA. The promoter in such an expression cassette can be constitutive, cell type-specific, stage-specific, and/or modulatable (e.g., by hormones such as glucocorticoids; MMTV promoter). Transcription can be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting sequences of between 10 to 300 base pairs that increase transcription by a promoter. Enhancers can effectively increase transcription when either 5' or 3' to the transcription unit. They are also effective if located within an intron or within the coding sequence itself. Typically, viral enhancers are used, including SV40 enhancers, cytomegalovirus enhancers, polyoma enhancers, and adenovirus enhancers.

Enhancer sequences from mammalian systems are also commonly used, such as the mouse immunoglobulin heavy chain enhancer.

Gene therapy vectors of all kinds can also include a selectable marker gene. Examples of suitable markers include, the dihydrofolate reductase gene (DHFR), the thymidine kinase gene (TK), or prokaryotic genes conferring drug resistance, *gpt* (xanthine-guanine phosphoribosyltransferase, which can be selected for with mycophenolic acid; *neo* (neomycin phosphotransferase), which can be selected for with G418, hygromycin, or puromycin; and DHFR (dihydrofolate reductase), which can be selected for with methotrexate (Mulligan & Berg, *Proc. Natl. Acad. Sci. (U.S.A.)* 78, 2072 (1981); Southern & Berg, *J. Mol. Appl. Genet.* 1, 327 (1982)).

efore integration, the vector has to cross many barriers which can result in only a very minor fraction of the DNA ever being expressed. Limitations to high level gene expression include: loss of vector due to nucleases present in blood and tissues; inefficient entry of DNA into a cell; inefficient entry of DNA into the nucleus of the cell and preference of DNA for other compartments; lack of DNA stability in the nucleus (factor limiting nuclear stability may differ from those affecting other cellular and extracellular compartments), efficiency of integration into the chromosome; and site of integration.

These potential losses of efficiency can be addressed by including additional sequences in a nonviral vector besides the expression cassette from which the product effecting therapy is to be expressed. The additional sequences can have roles in conferring stability both outside and within a cell, mediating entry into a cell, mediating entry into the nucleus of a cell and mediating integration within nuclear DNA. For example, aptamer-like DNA structures, or other protein binding sites can be used to mediate binding of a vector to cell surface receptors or to serum proteins that bind to a receptor thereby increasing the efficiency of DNA transfer into the cell.

Other DNA sequences can directly or indirectly result in avoidance of certain compartments and preference for other compartments, from which escape or entry into the nucleus is more efficient. Other DNA sites and structures directly or indirectly bind to receptors in the nuclear membrane or to other proteins that go into the nucleus, thereby facilitating nuclear uptake of a vector. Other DNA sequences directly or indirectly affect the efficiency of integration. For integration by homologous recombination, important factors are the degree and length of homology to chromosomal sequences, as well as the frequency of

such sequences in the genome (e.g., alu repeats). The specific sequence mediating homologous recombination is also important, since integration occurs much more easily in transcriptionally active DNA. Methods and materials for constructing homologous targeting constructs are described by *e.g.*, Mansour (1988) *Nature* 336:348; Bradley (1992) *Bio/Technology* 10:534.

For nonhomologous, illegitimate and site-specific recombination, recombination is mediated by specific sites on the therapy vector which interact with cell encoded recombination proteins, *e.g.*, Cre/Lox and Flp/Frt systems, as discussed above for *in vitro* systems. See also Baubonis (1993) *Nucleic Acids Res.* 21:2025-2029, which reports that a vector including a LoxP site becomes integrated at a LoxP site in chromosomal DNA in the presence of Cre recombinase enzyme.

Nonviral vectors encoding products useful in gene therapy can be introduced into an animal by means such as lipofection, biolistics, virosomes, liposomes, immunoliposomes, polycation:nucleic acid conjugates, naked DNA, artificial virions, agent-enhanced uptake of DNA, *ex vivo* transduction. Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024.

Unlike existing viral-based gene therapy vectors which can only incorporate a relatively small non-viral polynucleotide sequence into the viral genome because of size limitations for packaging virion particles, naked DNA or lipofection complexes can be used to transfer large (e.g., 50-5,000 kb) exogenous polynucleotides into cells. This property of nonviral vectors is particularly advantageous since many genes which can be delivered by therapy span over 100 kilobases (e.g., amyloid precursor protein (APP) gene, Huntington's chorea gene) and large homologous targeting constructs or transgenes can be required for efficient integration. Optionally, such large genes can be delivered to target cells as two or more fragments and reconstructed by homologous recombination within a cell (see WO 92/03917).

C. Applications of Gene Therapy

Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal,

intramuscular, subdermal, or intracranial infusion) or topical application. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

An important application is the treatment of congenital disease, particularly in patients lacking both wildtype alleles of a recessive gene. The vector introduces a wildtype allele of the gene that allows synthesis of the corresponding gene product compensating for the absence of this product in the patient. Examples of recessive diseases include sickle cell anemia, beta-thalassemia, phenylketonuria, galactosemia, Wilson's disease, hemochromatosis, severe combined immunodeficiency, alpha-1-antitrypsin deficiency, albinism, alkaptonuria, lysosomal storage diseases, Ehlers-Danlos syndrome, hemophilia, agammaglobulimenia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease and fragile X-syndrome.

Another application of gene therapy is to introduce a gene that increases the resistance of a cell to infection by pathogenic organisms. The gene can encode an antisense RNA to a sequence in the microorganism not found in the patient's genome. Alternatively, the gene can encode a protein inhibitory to the microorganism. Examples of microorganisms that can be inhibited by gene therapy include viral diseases (e.g., hepatitis (A, B, or C), herpes virus (*e.g.*, VZV, HSV-1, HAV-6, HSV-II, CMV, and EBV), HIV, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, cornovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, molluscum virus, poliovirus, rabies virus, JC virus and arboviral encephalitis virus) and pathogenic bacteria (e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumonococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lymes disease bacteria). For example, the HIV sequences Tat and Rev (Malim et al., *Nature* 338, 254 (1989)) are suitable targets for antisense RNAs or RNA binding proteins.

A further application of gene therapy in the delivery of drug resistance genes (polynucleotides conferring resistance to chemotherapeutic agents) to noncancerous cells in a patient with a view to increasing selective toxicity of the drug for cancer cells in the patient.

For example, polynucleotides conferring resistance to a chemotherapeutic agent (e.g., an expression cassette driving constitutive expression of the *hALDH-1* or *hALDH-2* gene can confer resistance to cyclophosphamide) can be transferred to non-neoplastic cells, especially hematopoietic cells. Other polynucleotides conferring resistance to chemotherapeutic agents include the cDNAs for ATP Binding Cassette transporters such as MDR1, MRP1, cMOAT, MRP3, MRP4, MRP5 (see EP-A-96200460 filed February 22 1996).

A further application of gene therapy is to infect CD34+ cells containing the hematopoietic stem cell and select for those cells expressing a drug resistance gene such as MDR1,MRP1,cMOAT, MRP3, MRP4, MRP5.

In another application, gene therapy vectors are used to deliver a negative selection gene to cells of a patient for which selective elimination is desired (e.g., cancer cells or cells of a pathogen). Examples of negative selection genes include ricin or diphtheria toxin, and HSV thymidine kinase (tk). Vectors bearing such genes can be selectively introduced into target cells via a cell surface receptor for which the vector has specific affinity. Expression of the negative selection gene (in the case of HSV tk in the presence of ganciclovir) kills cells bearing the gene.

In another application, gene therapy vectors can be used as vaccines to confer protection in subjects at risk of infection or to treat subjects who have already been infected. Such vectors encode immunogenic epitope(s) of pathogenic microorganisms and express the epitopes in the patient, particularly in target tissues at primary risk of infection, such as the oral and genital mucosa.

III. Applications of Recursive Sequence Recombination to Gene Therapy

The methods of the invention can be used to develop or improve on methods and materials used in gene therapy, including animals, cells and vectors for use in *in vivo*, *ex vivo* and *in vitro* systems. This section discusses the application of recursive sequence recombination to some specific goals in gene therapy. Many of these goals relate to improvements in vectors used in gene therapy. Unless otherwise indicated the methods are applicable to both viral and nonviral vectors.

(A) Improved Titer of a Viral Vector

In one embodiment, viruses with improved titers can be developed using the recursive recombination methods of the invention. The property of high viral titer can be an advantage in propagating large amounts of a virus *in vitro* for use as an agent in gene therapy. This property is also useful if it is desired that the virus replicate in a host tissue, such that progeny viruses infect cells surrounding the initially infected cell. Titer of a virus can be improved by recursive sequence recombination. The initial substrates for recombination can be viral genomes showing sequence divergence as a result of natural or induced variation. The substrates can be whole genomes or fragments thereof. Recombination of fragments is useful for large genomes or in situations in which a part of the viral genome is known to be particularly important in conferring high titer. The substrates can be recombined *in vitro* or can be introduced into cells and recombined *in vivo*. Recombination *in vivo* can be used to generate progeny viruses that can be screened directly. However, recombination *in vitro* leads to recombinant genomes or fragments thereof. Whole recombinant genomes can be packaged into viruses using a packaging cell line or an *in vitro* packaging system. Fragments of genomes are usually first assembled by DNA ligation. They are subsequently inserted into a viral genome before packaging. Irrespective of the precise route, one arrives at a population of viruses having genomes at least part of which constitutes a recombinant segment.

The collection of viruses with recombinant genomes can be screened simply by propagating the viruses in cell culture for several generations. The viruses with the highest titer thereby acquire the highest representation among progeny viruses. If desired, viruses can be plaque-purified and titers of individual viruses compared to identify the very best titer of viruses from a round of recombination. Alternatively, the viruses can be purified by serial dilution to determine the very best titer viruses from a round of recombination.

The genomes from viruses surviving screening are subject to a further round of recombination, which again can be performed *in vivo* or *in vitro*. For *in vivo* recombination, viruses having genomes containing the recombinant segments can, for example, be infected into a cell at high multiplicity. For *in vitro* recombination, viral DNA is isolated from viruses harboring recombinant DNA. The genomes from viruses surviving screening can be recombined with each other or with fresh substrates obtained from similar sources to the initial substrates. In some recombination steps, it is desirable to include an excess of

wildtype version of the viral genome to reduce silent mutations. Again, recombination can be performed with whole genomes or fragments thereof. Selection is repeated as before.

After several rounds of recombination and selection, viral mutants, or clones, capable of producing the desirable titer can be obtained. For example, without concentration of an infected cell culture, it is possible to achieve a concentration of evolved virus of at least 10^6, 10^8 or 10^{10} viruses/ml.

(B) Improved Infectivity of a Virus

The infectivity of a virus means the percentage of viruses that infect a cell when an inoculum of viruses is contacted with an excess of cells. Obtaining a high infectivity is particularly important with respect to the intended target cell-type. Thus, if a viral vector is being used to deliver a beneficial expression product to a target tissue (e.g., lung cells lacking a functional endogenous CFTR gene), it is usually desirable that as high a percentage of viruses as possible infect that cell type.

The selection of substrates and means of recombination follows the same principles as discussed for improved viral titer. However, the means of screening viruses bearing recombinant genomes is usually different. The previous selection does not necessarily select for viruses having high infectivity because high titer can also be conferred by high burst size per cell. To screen more specifically for high infectivity, clonal isolates of viruses bearing recombinant segment are used to infect separate cultures of cells. The percentage of viruses infecting cells can then be determined by, for example, counting cells expressing a marker expressed by the viruses in the course of infection. After several rounds of recombination and screening, viruses harboring recombinant genomes capable of infecting 50, 75, 95 or 99% of target cells are obtained.

(C) Improved Packaging Capacity of a Virus

Viruses and vectors with the capability of incorporating increasing amounts of recombinant nucleic acids sequences, such as having an improved packaging capacity within the viral capsid, can be developed using the recursive recombination methods of the invention. As noted above, the viruses commonly used in gene therapy can package only a limited genome length, thus, restricting the capacity of viruses to accommodate large inserts. Capacity of a virus can be improved using similar principles to those discussed above. In these methods, the viral genome to be lengthened should have a site into which increasing lengths of nucleic acid can be inserted in successive rounds of screening without affecting

other viral functions. Initially, one can start with a viral genome having an insert such that the combined length of the genome is close to the existing maximum capacity of the virus. The initial substrates for recombination are variant viral genomes as in the other methods. The variation usually occurs other than in the length-conferring insert because the insert is replaced in actual use of the vector. One source of starting substrates can be viral genomes known to show sequence similarity with the virus to be evolved but which have a larger genome packaging capacity. Recombination proceeds in the same manner as discussed above. Viruses having recombinant genomes are then screened for titer or infectivity as discussed above. Recombinant genomes from viruses having the best titer and/or infectivity are manipulated to introduce a further insert to increase the genome length. There follow further cycles of recombination, screening and increasing genome length, until viruses are achieved that can accommodate inserts of the desired size. For example, the maximum insert size used in most existing adenoassociated viral vectors is about 5 kb, which can be increased to 10, 15, 20 or 50 kb or more.

(D) Improved Stability of a Virus

Viruses with improved stability can be developed using the recursive recombination methods of the invention. Stability of a virus for use in gene therapy is important both in prolonging the shelf-life of the virus as a drug between manufacture and administration, and in the subsequent ability of the virus to resist cellular degradative mechanisms before reaching its target. The principles for selection of starting substrates and performing recombination are the same as in other methods described above. Viruses bearing recombinant genomes that have evolved to acquire greater stability can be selected by exposing the viruses to destabilizing conditions and recovering surviving viruses. For example, destabilizing conditions include temperature (hot or cold), mechanical disruption (e.g., centrifugation or sonication), exposure to chemicals or exposure to biological degrading agents such as proteases (e.g. serum proteases). Viruses surviving exposure to destabilizing conditions are identified by propagation of treated viruses and collection of progeny. Sometimes, propagation proceeds only for one or a limited number of generations, since otherwise progeny viruses become biased toward those having genomes favoring high titer in addition to those having genomes conferring stability.

(E) Improved Expression or Expression Regulation of a Vector Coded Sequence

Improved expression of a gene sequence of interest can be achieved by performing the recursive sequence recombination methods of the invention. Usually viral or nonviral vectors used in gene therapy encode a product to be expressed in an intended target cell. The product can be a protein or RNA, such as an antisense RNA or RNA that specifically binds a target protein, *i.e.*, an aptamer. Usually, the coding sequence is operably linked to an additional sequence, such as a regulatory sequence, to ensure its expression, such as some or all of the following: an enhancer, a promoter, a signal peptide sequence, an intron and/or a polyadenylation sequence. A desirable goal is to increase the level of expression of functional expression product relative to that achieved with conventional vectors. Expression can effectively be improved by a variety of means, including increasing the rate of production of an expression product, decreasing the rate of degradation of the expression product or improving the capacity of the expression product to perform its intended function. Improvement of the latter four parameters for drug transporters including but not limited to MDR1, MRP1, cMOAT, MRP3, MRP4, MRP5, an embodiment of this invention, results in preferred variants of these transporters. These are applied in protective gene therapy of a wide variety of tissues including but not limited to bone marrow, kidney, liver, intestine and heart. These improved drug transporter variants are also applied in dual vectors such as dual retroviral vectors which carry the transporter variant and a gene encoding a therapeutic gene such as the gene for lysosomal glucocerebrosidase deficient in Gaucher disease. *In vivo* selection for the improved drug transporter variant present on the dual construct results in selection for the therapeutic sequence as well and thus has therapeutic benefit.

Improved expression of selection markers can be achieved by performing recursive sequence recombination. For purposes of selection, a gene product expressed from a vector is sometimes an easily detected marker rather than a product having an actual therapeutic purpose, *e.g.*, a green fluorescent protein (see Crameri (1996) *Nature Biotech.* 14:315-319) or a cell surface protein. However, some genes having a therapeutic purpose, e.g., drug resistance genes, themselves provide a selectable marker, and no additional or substitute marker is required. Alternatively, the gene product can be a fusion protein comprising any combination of detection and selection markers.

The substrates for recombination can be the full-length vectors or fragments thereof including coding sequence and/or regulatory sequences to which the coding sequence

is operably linked. The substrates can include variants of any of the regulatory and/or coding sequence(s) present in the vector. If recombination is effected at the level of fragments, the recombinant segments should be reinserted into vectors before screening. If recombination proceeds *in vitro*, vectors containing the recombinant segments are usually introduced into cells before screening. Cells containing the recombinant segments can be screened by detecting expression of the gene encoded by the selection marker. Internal reference marker genes can be included on the vector to detect and compensate for variations in copy number or insertion site. For example, if this marker is green fluorescent protein, cells with the highest expression levels can be identified by FACS™. If the marker is a cell surface protein, such as MDR1 or cMOAT, the cells are stained with a reagent having affinity for the protein, such as antibody, and again analyzed by FACS™. Recombinant segments from the cells showing highest expression are used as some or all of the substrates in the next round of screening.

Evolution of Cytomegalovirus Transcriptional Regulatory Elements

The major immediate-early (IE) region transcriptional regulatory elements, including promoter and enhancer sequences (the promoter/enhancer region), of cytomegalovirus (CMV) is widely used for regulating transcription in vectors used for gene therapy because it is highly active in a broad range of cell types. Optimized CMV transcriptional regulatory elements which direct increased levels of transgene expression is generated by the recursive recombination methods of the invention, resulting in improved efficacy of gene therapy. As the CMV promoter and enhancer is active in human and animal cells, the improved CMV promoter/enhancer elements are used to express foreign genes both in animal models and in clinical applications.

A library of chimeric transcriptional regulatory elements is created by DNA shuffling of wild-type sequences from five related strains of CMV. The promoter, enhancer and first intron sequences of the IE region are obtained by PCR from the CMV strains: human VR-538 strain AD169 (Rowe (1956) *Proc. Soc. Exp. Biol. Med.* 92:418; human V-977 strain Towne (Plotkin (1975) *Infect. Immunol.* 12:521-527); rhesus VR-677 strain 68-1 (Asher (1969) *Bacteriol. Proc.* 269:91); vervet VR-706 strain CSG (Black (1963) *Proc. Soc. Exp. Biol. Med.* 112:601); and, squirrel monkey VR-1398 strain SqSHV (Rangan (1980) Lab. Animal Sci. 30:532). The promoter/ enhancer sequences of the human CMV strains are 95% homologous, and share 70% homology with the sequences of the monkey isolates, allowing

the use of family shuffling to generate a library great diversity. Following shuffling, the library is cloned into a plasmid backbone and used to direct transcription of a marker gene in mammalian cells. An internal marker under the control of a native promoter can be included in the plasmid vector. Expression markers, such as green fluorescent protein (GFP) and CD86 (also known as B7.2, see Freeman (1993) *J. Exp. Med.* 178:2185, Chen (1994) *J. Immunol.* 152:4929) can also be used. In addition, transfection of SV40 T antigen-transformed cells can be used to amplify a vector which contains an SV40 origin of replication. The transfected cells are screened by FACS sorting to identify those which express high levels of the marker gene, normalized against the internal marker to account for differences in vector copy numbers per cell. If desired, vectors carrying optimal, recursively recombined promoter sequences are recovered and subjected to further cycles of shuffling and selection.

(F) Improved Expression and/or Function of Drug Resistance Sequences

The recursive recombination methods of the invention also provide for means to improve the expression of drug resistance sequences/ proteins. Many treatment regimes entail administration of drugs having side-effects on a particular cell type in the body. For example, chemotherapy is notorious for killing cells other than the targeted cancer cells. See Licht (1995) *Cytokines & Molecular Therapy* 1:11-20. Myelosuppression, or bone marrow toxicity, is dose-limiting for many chemotherapeutic agents. This is not only a dangerous side effect but also limits the effectiveness of chemotherapy. Indeed, the chemotherapy can be fatal, either directly by loss of blood cell function or indirectly by causing secondary cancers such as leukemia. It is possible to protect hematopoietic cells by delivering drug resistance proteins via gene therapy. This principle has been demonstrated by a number of studies in which murine bone marrow cells were protected against chemotherapeutic alkylating agents by the overexpression of a protective alkyltransferase. Other drug resistance proteins can be used for chemoprotection of normal tissues and can be targets for improved expression using the methods of the invention. They include, for example, glutathione-S-transferase, dihydrofolate reductase and superoxide dismutase.

Alkylating agents are especially toxic to the hematopoietic system, with myelosuppression being the dose-limiting side effect. Hematopoietic cells are so susceptible to alkylating agents that iatrogenic leukemias are a common occurrence. Alkylation therapy can also cause severe pulmonary toxicity and result in dose limitations. Examples of other

drugs that have dose limitations due to toxicity to vital organs are etoposide (e.g. kidney), cisplatin (e.g. kidney), taxol (e.g. lung), anthracyclines (e.g. heart), See Perry et al, The Chemotherapy Source Book,1991, Williams and Wilkins, Baltimore, USA, ISBN 0-683-06859-08. This limitation sensitivity can be attributed to the low expression of the DNA repair protein O^6-methylguanine-DNA methyltransferase in hematopoietic cells (also called O^6-alkylguanine-DNA alkyltransferase, MGMT or alkyltransferase; EC 2.1.1.63). Alkylating agents, especially nitrosoureas, as used either alone or in combination with other drugs to treat many types of cancer, such as Hodgkin's and non-Hodgkins lymphomas, multiple myeloma, malignant melanoma, brain neoplasms, gastrointestinal cancers and lung cancers. Together these cases constitute over one third of all cancers diagnosed. Thus, improving the effectiveness and decreasing the toxicity of alkylation-based chemotherapeutic regimens will have a profound impact on health care.

The introduction of drug-resistance genes into bone marrow stem cells or pulmonary cells or kidney cells or heart cells or liver cells or intestinal cells via gene therapy is one way to overcome the limitations of chemotherapeutic regimens. In the case of bone marrow, one strategy is to transduce the cells *ex vivo* with the drug resistance gene and repopulate the bone marrow with these cells before or after chemotherapy. Bone marrow is a relatively easy tissue to extract, manipulate and reintroduce into the body. Kidney or liver or heart or intestine or central nervous tissue or other tissues are protected by retrovirus or adenovirus or AAV vectors or nonviral vectors carrying drug resistance genes after *in vivo* administration of the recombinant adenovirus into the patient and targeting of the virus to the desired tissue followed by chemotherapy aimed at the killing of a tumor in a tissue other than the protected tissue.

MGMT is found in all organisms examined, prevents the mutagenic, cytotoxic, and carcinogenic effects of chemotherapeutic alkylating agents. MGMT removes alkyl groups attached by such chemicals from the O^6 position of guanine. These alkyl groups are transferred irreversibly to a cysteine in the active site of the MGMT protein, inactivating the alkyltransferase. Thus, the enzyme is a suicide enzyme and can act only stoichiometrically, which is an important barrier to improvement of MGMT. Because each protein module acts only once in a suicidal manner, the protection afforded a cell is determined not only by the activity (quality) of the MGMT but also by the number of MGMT molecules. Cells, such as bone marrow cells, which express little or no alkyltransferase are very sensitive to laboratory

alkylating agents such as N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) (Day (1980) *Nature* 288:724-727) and clinically used nitrosoureas (Erickson (1980) *Nature* 288:727-729). Thus, myelosuppression is a serious problem with drug-based chemotherapeutic regimens (DeVita (1993) *Cancer: Principles and Practice of Oncology*), but it has been overcome in experiments in which the wild-type human, mouse, or bacterial alkyltransferase genes were transduced into human and mouse hematopoietic cells. The overexpressed genes, carried on retroviral vectors, protected stem cells in culture from killing by nitrosoureas (Allay (1995) *Blood* 85:3342-3351; Moritz (1995) *Cancer Res.*55:2608-2614). Furthermore, when these cells were transplanted into the bone marrow of mice, the protection proved to be long-lasting *in vivo* (Maze (1996) *Proc. Natl. Acad. Sci. USA* 93:206-210). Similar effects were seen when liver and thymus rather than bone marrow were targeted (Dumenco (1993) *Science* 259:219-222; and Nakatsuru (1993) *Proc. Natl. Acad. Sci.USA* 90:6468-6472).

This protective effect of MGMT can be improved by recursive sequence recombination in several respects. First, novel variants can be selected having higher specific activity, *i.e.*, faster repair of cytotoxic alkylation-induced lesions. Thus, for a given expression level, bone marrow cells will be better protected. Some improvement in MGMT has been reported (Christians (1996) *Proc. Natl. Acad. Sci. USA* 93:6124-6128) using a conventional cassette mutagenesis. Second, novel variants can be selected for resistance to inhibitors of wild-type alkyltransferases, such as O^6-benzylguanine. Such inhibitors are sometimes used to suppress endogenous alkyltransferases present in cancer cells (Pegg (1995) *Progress in Nucleic Acid Res. and Molec. Biol.* 51:167-223). Inhibitor-resistant MGMT can be used to transfect bone marrow in treatment protocols in which alkylating agents are combined with inhibitors of alkyltransferases. Third, novel variants of the coding sequence and/or operably linked regulatory sequences can be selected for improved expression of MGMT. Fourth, variants of MGMT can be produced that bind to but do not remove alkyl adducts from DNA, effectively resulting in DNA-protein crosslinks more toxic to the cell than the alkyl adducts alone. Vectors expressing the mutant variants can be targeted to cancer cells before treatment with the alkylating substrate. Fifth, MGMT variants can be selected to protect mammalian cells against the clinically relevant nitrosoureas. For this purpose, selection should be preferably performed in mammalian cells rather than bacterial cells, because the protective effect of MGMT against nitrosoureas is stronger in the former.

The sometimes-low transfection efficiencies of gene therapy are not a major limitation in *ex vivo* methods because alkylation treatment effectively serves as a positive selection for transfected cells. In contrast, low transfection efficiencies can be a problem in *in vivo* gene replacement therapy because there is no generally positive selection, only negative selection by tumoricidal gene therapy. Improved means of positive selection for *in vivo* gene replacement therapy allows, for example, a relatively small number of chemoresistant hematopoietic cells to repopulate the bone marrow.

A drug-resistance gene is a starting material for improvement using the methods of the invention is the multi-drug resistance gene *MDR*-1. *MDR*-1 encodes a plasma membrane glycoprotein called "P-glycoprotein (Pgp") which acts as an ATP-dependent drug efflux pump and confers chemoresistance to a wide variety of drugs (Chin (1993) *Adv. Cancer Res.* 60:157). Cells not expressing *MDR*-1 are exquisitely sensitive to drugs such as vincristine, etoposide, and colchicine. This same chemoresistance property, which when expressed by tumor cells can frustrate chemotherapy efforts, can be turned to an advantage when used as a positive selectable marker. Metz (1996) *Virology* 217:230-241, reported a 20-fold higher stringency when selecting for *MDR*1 expression compared to *neo* selection. P-glycoprotein has been demonstrated to positively select for transformed cells in the *in vitro* correction of cells from at least two different genetic diseases, Fabry disease (Sugimoto (1995) *Human Gene Therapy* 6:905-915) and chronic granulomatous disease (Sokolic (1996) *Blood* 87:42-50). However, there is no reason to believe that nature has optimized *MDR*1 for activity against man-made drugs. Improving *MDR*-1 by recursive recombination to improve protection of cells from drugs such as etoposide and colchicine will allow the use of higher levels of such selective agents, which will increase the selection stringency and better differentiate between transformed and non-transformed cells.

MDR-1 is improved/modified by DNA shuffling followed by positive selection in mammalian cells. Randomly mutated pools of *MDR*-1 are inserted into appropriate vectors (*e.g.*, retroviral, adenoviral vectors) and transformed into drug-sensitive cells. Selection with colchicine and/or etoposide and/or vincristine will identify active *MDR*-1 variants. The *MDR*-1 genes are rescued from surviving cells and subjected to additional rounds of recombination and selection with increasing doses of drugs.

Because some mammalian cells already express high levels of P-glycoprotein, it might be difficult to determine whether the improved *MDR*-1 transgene is expressed in

these cells; *i.e.*, the background will be high. In this case the endogenous P-glycoprotein is inactivated with a well-characterized inhibitor such as verapamil, and transform with a marker *MDR*-1 transgene that encodes a mutant P-glycoprotein resistant to the inhibitor yet highly active against the cytotoxic drug. Such a variant is created by selecting *MDR*-1 mutant pools in the presence of both the inhibitor and the cytotoxic drug(s), such as colchicine. For example, the methods of the invention are used to create alkyltransferase mutants super-active against the cytotoxic chemical N-methyl- N-nitro- N-nitrosoguanidine (MNNG) and completely resistant to the alkyltransferase inhibitor benzylguanine.

MDR-1 thus optimized as a positive selection marker is inserted into the vector of choice. The vector can also be optimized by DNA shuffling, either by itself or in combination with *MDR*1 mutagenesis (*MDR*1 and the vector shuffled as a unit). Shuffling the entire construct allows many parameters to be tested at once. Bicistronic arrays, 2 genes transcribed as one mRNA from the same promoter but translated from separate ribosome binding sites, can be used (Sugimoto (1995) *Human Gene Ther., supra*). Shuffling the entire array or the whole construct can be used to optimize secondary structure of the bicistronic mRNA to improve translation of the second, downstream gene. For example, a bicistronic retroviral vector encoding *MDR*1 and a gene complementing a genetic defect can be constructed and optimized using the methods of the invention. The entire vector can be mutagenized by DNA shuffling and reassembled. Additionally, the vector can be packaged as virus by a packaging cell line, transfected into the defective cells, and selected with colchicine. Selection is effected by analyzing surviving cells for complementation of the genetic defect.

Further candidates for improvement are members of the ATP Binding Cassette (ABC) family of transporters. Members of this family include but are not limited to MDR1, MDR2, MRP1. MDR1 and MRP1 encode ATP dependent drug efflux pumps useful for the protection of stem cells in an *ex vivo* gene therapy setting. Other ABC transporters include the canalicular Multispecific Organic Anion Transporter (cMOAT), MRP3, MRP4 and MRP5 subject of patent application EP-A-96200460 (filed February 22 1996). cMOAT is involved in the transport of organic anions such as glutathione and glucuronide conjugates of cis-platinum and etoposide of which the parent compounds are used in cancer treatment regimens (Paulusma (1996) *Science* 271:1126-1128). Desired chemotherapeutic agents such as etoposide and mitoxantrones do not represent good substrates for MDR1 or cMOAT but

are drugs that are clinically very desirable agents and therefore mutant versions transporting etoposide, mitoxantrones or cisplatin with high efficiency are useful for protective gene therapies including gene therapies using MDR1 and cMOAT.

A drug transporter gene can be evolved/modified not only to confer improved protection to drugs it already recognizes (*e.g.*, etoposide) but also to confer protection against drugs not recognized by wildtype *MDR*-1, such as alkylating agents. For example, an *MDR*-1 gene can be modified by recursive recombination (evolved) to pump alkylating agent out of a cell, thus serving as a complement to MGMT (described above). For example, both the MGMT and *MDR*-1 genes can be transduced into stem cells before combination chemotherapy in which one of the drugs is an alkylator. Studies in which stem cells were transduced with the wild-type MDR-1 gene gave results similar to those cited above with MGMT for alkylating agents (Sorrentino (1992) *Science* 257:99).

Another suitable gene for evolution/ modification using the methods of the invention is glutathione-S-transferase, which detoxifies alkylating drugs in the cytoplasm, complementing MGMT. It acts on drugs after they have entered the nucleus and damaged DNA. Some improvements in glutathione-S-transferase resulting from conventional cassette mutagenesis in bacteria have already been reported (Gulick (1995) *Prod. Natl. Acad. Sci. USA* 92:8140-8144). Further evolution by recursive sequence recombination will provide additional improvements. The improvement gene can then be transfected into stem cells or lung cells on its own or in combination with MGMT.

Other drug-resistance genes are candidates for evolution for use in suppressing side effects in other tissues. For example, bleomycin is an antineoplastic whose major toxicity is to pulmonary cells. The protein bleomycin hydrolase can protect cells from bleomycin, and the human gene was recently cloned (Bromme (1996) *Biochemistry* 35:6706-714). The gene can be improved by gene shuffling and used to protect pulmonary cells in cancer patients.

Inhibition of replication and spread of infective HIV-1 by retroviruses expressing anti-HIV molecules such as HIV specific antisense or ribozymes have been shown to be a promising approach for the treatment of HIV-1 infected individuals. Such therapy can only be expected to be successful in the long run, when virus replication is prevented in the majority of CD4+ (HIV-1 permissive) cells. Most of the CD4+ T-lymphocytes and macrophages have a limited lifespan so that transduction of these cells can provide no lasting

protection. Therefore, hematopoietic stem cells are the target cells of choice for HIV gene therapy. Unfortunately, preclinical and clinical studies demonstrate that after retransplantation of transduced hemopoietic stem cells only 0.1% of the peripheral blood cells contain the virus. Introduction of a gene that enables *in vivo* selection of transduced cells next to the antiviral polynucleotide sequence may overcome this problem. In another application MDR1 or MRP1 or cMOAT or MRP3 or MRP4 or MRP5 variants are generated that more efficiently pump HIV inhibitors such as the clinically used reverse transcriptase inhibitors AZT and ddC or HIV protease inhibitors or combinations thereof. These are desired for use in stem cell based anti HIV gene therapy using *in vivo* selection of AZT resistant stem cells carrying an AZT transporting MDR1 or cMOAT variant and an anti-HIV sequence such as a ribozyme or antisense sequence. Since AZT and ddC are known for their toxic effects on hematopoietic cells, the MDR1/AZT system provides an efficient *in vivo* selection system for stem cell-based gene therapy protocols to treat HIV infected individuals.

In other embodiments, candidate genes for improvement include the genes encoding DNA ligase and topoisomerase to protect against ionizing radiation (Boothman (1994) *Cancer Res.* 54:4618-4626), genes encoding nucleotide excision repair enzymes such as T4 endonuclease V to protect against UV irradiation and skin cancer, and genes encoding alkaline phosphatase endonuclease and glycosylases to improve the base excision repair pathway which is crucial to ward off the effects of oxidative DNA lesions thought to cause many types of cancer and accelerated aging.

Evolution/modification of drug-resistance genes and associated regulatory sequences using the methods of the invention falls under the general approach discussed above for improving gene expression. However, in evolving drug-resistance genes, it is sometimes desirable not only to improve expression of the gene but to increase the degree of resistance conferred by the gene product with a particular drug. In this situation, it is preferable that substrates for recombination include the drug-resistance gene as well as associated regulatory sequences so that the resistance gene can be evolved within the genetic context in which it is to be expressed. Diversity between the initial substrates can be the result of induced mutations, natural drug-resistance genes from different sources, and mutations already known to confer improved properties.

For example, the cDNA sequences of five different mammalian species of MGMT (human, rat, mouse, hamster, and rabbit) have been reported, and, despite very

extensive homology, variations do exist, as illustrated in Figure 4. Following is an alignment showing the human amino acid sequence on the top line with other amino acid sequence found in nature shown below the human sequence.

MDKDCEMKRT TLDSPLGKLE LSGCEQGLHE IKLLGKGTSA ADAVEVPAPA 50
AET KL YS VFH AM R G RFPSGK PN T PT A TP
I D K A I S SS K C
E S

AVLGGPEPLM QCTAWLNAYF HQPEAIEEEP VPALHHPVEQ QESFTRQVLW 100
EL S V ET E RE A TPGL L E
G H Q S

KLLKVVKFGE VISYQQLAAL AGNPKAARAV GGAMRGNPVP ILIPCHRVVC 150
TV S IR
M N

SSGAVGNYSG GLAVKEWLLA HEGHRLGKPG LGGSSGLAGA WLKGAGATSG 200
D SI H Q IPTRQ A SKGL I S R SSFESTS
N A C D TT P G

SPPAGRN 207
PELS
K

The natural variations can be incorporated by any of the formats discussed in Section II to generate recombinant forms of MGMT including natural segments unique to human and nonhuman forms, as discussed in Example 3. For example, oligonucleotides can be designed to encode all the different combinations of natural variants, and these oligonucleotides will be mixed in with the fragmented wild-type human gene. A surprisingly small number of oligonucleotides (twenty-one) can be used if they are degenerate at positions at which more than two amino acids are represented in nature (see Figure 3). The oligonucleotides shown in Figure 3 contain up to twenty one bases of nonhomology to the human sequence flanked on either side by a 20 base sequence perfectly matched with the human MGMT sequence. Another use of "oligo spiking" is to bias shuffled gene pools toward known desirable mutations such as the V139F mutation demonstrated to improve the wild-type protein (Christians (1996) *Prod. Natl. Acad. Sci. USA* 93, 6124-6128), or mutations conferring O^6-benzylguanine resistance.

An alternative to "oligo spiking" is to obtain all the individual cDNAs and shuffle them together. This option might have some tendency to dilute the human character of the pool leading to immunogenic problems when used in human gene therapy, but this problem can be overcome by backcrossing mutants with the wild-type human gene to eliminate non-useful mutations.

Recombined drug-resistance genes and vectors encoding them can readily be screened for improved expression. Cells containing vectors containing recombinant segments are exposed to the drug and surviving cells recovered. These cells are enriched for recombinant segments conferring improved resistance to the drug. Screening can be made more stringent in successive rounds by increasing the concentration of drug or duration of exposure thereto.

The final round of selection is usually performed in stem cells because some of the component factor contributing to the end point of drug-resistance may be cell-type dependent (see Examples 5 and 6). Because expression levels are important for the protective effect, manipulating vector sequences other than that encoding drug resistance genes such as MGMT, MDR1, cMOAT, MRP1, MRP3, MRP4 and MRP5, provides an important source of improvement. The vectors are selected based on desired endpoints, such as the ability to protect cells from alkylating agents. The endpoint is achieved by a variety and a combination of components too complicated to predict, including enhanced transduction, better vector stability, and improved transcription of the gene in addition to improved or altered function of the drug resistance gene.

(G) Evolution of Transducing Vectors for Integration and Stable Expression in Mammalian Stem Cells

Vectors having new and/or improved ability to infect, integrate and express themselves in hematopoietic stem cells can be developed using the recursive recombination methods of the invention. A major goal in gene therapy is to develop practical methods to efficiently integrate DNA constructs into human stem cells. A practical method for efficiently integrating retroviruses into stem cells allows repopulation of patients with autologous bone marrow that had been genetically modified with traits of interest. For example, the stem cells are engineered to express trans-dominant factors that interfere with viral replication. Stem cells are engineered to express wild type or engineered transgenes that complement a defined genetic defect, such as Gaucher's disease. MDR genes or

alkyltransferase genes are inserted into stem cells to confer resistance to chemotherapeutic agents. Gene encoding T cell receptors specific for cancer or pathogen epitopes of interest are inserted for expression upon maturation of the stem cell.

However, stem cells are difficult to purify and rapidly lose their pluripotent phenotype if propagated *in vitro*. Retroviruses are very inefficient at integrating into nondividing cells in general, and stem cells in particular. Thus, recursive recombination is used to evolve a factor or set of factors that, upon infection with and expression of the retrovirus genes prior to integration, can transiently or permanently render a stem cell susceptible to retroviral integration while at the same time remaining pluripotent:

In one embodiment, large ($>10^6$) libraries of retroviruses expressing candidate factors for transiently perturbing stem cells so as to promote retroviral integration are made. Such factors include, but not be limited to: HIV matrix, HIV vpr, random fragments of HIV and other lentiviruses (the only class of retroviruses able to efficiently transduce non-dividing cells); cDNAs from stem cells; cDNA from stromal cell cultures (which make factors that influence the differentiation state of stem cells, and over production or evolution of recombined forms exert the desired effect); or, any other cytokine or growth factor. Such libraries are used in the *in vitro* and *in vivo* recursive recombination methods of the invention, as generally described above, to create a retrovirus which can efficiently infect, integrate and express sequences and proteins of interest in non-dividing stem cells.

Another embodiment repopulates SCID or SCID/NOD mice with human stem cells that have been transduced by a retrovirus modified by the above methods. Progeny of retroviruses from stem cells that were successfully transduced by a member of the initial retrovirus recombinant segment library are recovered. Selection markers, such as green fluorescent protein (GFP), drug markers, or cell sorter (FACS) markers may be encoded in the transducing retrovirus to facilitate recovery of repopulating stem cells transduced with a retrovirus construct. Sequences encoding the factors to be evolved/modified or the entire integrated retroviral genome can be recovered. Further rounds of recursive sequence recombination can be repeated until the desired efficiency/efficiency of stem cell transduction is achieved.

A murine SCID/NOD immunodeficient system that can be repopulated with primitive human hematopoietic stem cells can be used (Dick (1996) CSH Gene Therapy abstract #11). Retroviruses can infect these stem cells with very low but detectable

efficiency. Progenitor cells with integrated retrovirus can be recovered from peripheral blood cells in this SCID/NOD repopulation model. This and analogous repopulation systems therefore forms the basis for selecting retroviruses with improved efficiency of integration into primitive pluripotent cells. As noted above, including GFP in the vector allows for FACS purification of cells expressing retroviral-encoded proteins after repopulation. If the repopulation is initially very inefficient, a selectable gene such as Neo or TK to selectively culture transduced cells is also expressed.

In another embodiment, rather than removing infected stem cells and isolating retroviral sequences for further rounds of recursive recombination, lethally irradiated retroviral producing helper lines containing recombinant sequences are injected into the SCID/NOD bone marrow. With this technique, recursive recombination takes place *in vivo*; the stem cells remaining in the special environment of the bone marrow, an environment that may prove impossible to mimic *in vitro*.

In a further embodiment, recursive recombination is used to develop a means by which viruses which cannot normally lack the means to integrate into non-dividing cells. This method incorporates HIV proteins which are required for HIV to integrate into nondividing cells, into other vectors of interest. For example, integrase, the enzyme responsible for mediating the integration of the viral genome in the host cell chromosome, can suffice to connect the HIV-1 preintegration complex with the cell nuclear import machinery. Viral matrix and Vpr proteins also play important roles in the ability of HIV to integrate into non-dividing cells. See Gallay (1997) *Proc. Natl. Acad. Sci. USA* 94:9825-9830. Repeated cycles of recursive recombination, as DNA shuffling, are carried out until the desired property is conferred to the vector or sequence of interest.

In another embodiment, before recursive recombination, long term bone marrow cultures are stimulated to cycle *in vitro*. This results in increased retroviral transduction of the stem cells in both a murine SCID/Beige repopulation assay (Agatsuma (1997) ***Antiviral Res.*** 34:121-130) and in stem cell repopulation of terminal human myeloma patients with transduced bone marrow cells. Cycling stem cells are more susceptible to transduction. Thus, stem cells can be stimulated such that they are more susceptible to retroviral transduction and yet remain pluripotent.

(H) Improved Tissue Specificity of a Vector

Vectors with new and/or improved tissue specificity (tissue tropism) can be developed using the recursive recombination methods of the invention. In most gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Specificity of cellular targeting is a key issue impacting the safety and practicality of these vectors for *in vivo* gene therapy. Thus, there is a need to restrict and/or redirect the specificity of gene therapy vectors, such as adenovirus.

One example illustrating the need to deliver a gene therapy vector a specific tissue type involves delivering a wildtype CFTR gene to cystic fibrosis patients. The CFTR gene should be delivered mainly to pulmonary tissue. In a second example, where the gene therapy vector encodes a chemotherapeutic agent, it is desirable that the agent be delivered to neoplastic cells and not normal cells.

The strategy in selecting substrates and recombination formats is in general similar to those discussed before. Substrates for recombination can be whole viral genomes or can be fragments encoding the viral proteins thought to interact with cellular receptors. If such fragments are recombined, the recombination products should be reinserted into viral genomes, and the genomes packaged to form viruses before screening.

For example, for evolution of vesicular stomatitis virus (VSV) to infect new target cells, recursive recombination should focus on G-protein sequences, because the G protein is expressed on the capsid's outer surface (Schnell (1996) *Proc. Natl. Acad. Sci. USA* 93:11359-11365). Furthermore, it has been technically difficult to generate viruses encoding the vesicular stomatitis virus G-protein (VSV-G) because it is too toxic to the host cells to allow for viral propagation (Yoshida (1997) Biochem. Biophys. Res. Commun. 232:379-382). Thus, the methods of the invention can be used to generate modified VSV G protein, thereby generating new target cells for recombinant VSV.

There is also a need to generate tissue-specific adenoviruses. Since the tropism of adenovirus is nonselective, tissue-specific expression of systemically administered vectors can only be achieved by the use of a tissue-specific promoter/enhancer that is small enough to fit the insert capacity of the vector. Alternatively, tissue-specific expression is generated by ablating the native promiscuous tropism of adenovirus and constructing new tissue-specific domains using the methods of the invention. Generation of tissue-specific

adenoviruses by recursive sequence recombination overcomes this non-selective tropism limitation of native adenovirus in the use of the vector in gene therapy.

Adenovirus binds to eukaryotic cells using a "fiber protein" which protrudes from each of the twelve vertices of its icosahedral capsid. Serological and mutagenesis studies make it clear that the fiber, a homotrimer consisting of "staff" and "knob" domains, interacts with cellular receptors. The structure of the knob has been reported by Xia (1994) *Structure* 2:1259-1270. R. D. Gerard has used the structure of the heterotrimeric knob to scan this structure by SDM for mutations that reduce binding to the receptor (personal communication, 1996 CSH Gene Therapy meeting. These studies allow construction of mutants with abrogated or severely reduced ability to infect using the natural receptor, which is known to be expressed in many tissue types. This is a starting point from which to evolve, *i.e.*, use the recursive sequence recombination methods of the invention, new tissue specificities for the adenovirus fibers which bind to cellular receptors. V. Legrand (CSH poster 184) and Dan von Seggery (CSH poster #223) have reported systems for expressing mutants of the fiber protein off of a small easily manipulated SV40 based vector. These constructs will support plaque formation by an adenovirus deleted for the fiber gene. Legrand used this system to fuse the 11 amino acid Gastrin Releasing Peptide (GRP) to the C-terminus of the fiber gene. LacZ+ adenoviral mutants expressing this fusion protein were able to infect cells expressing GRP receptor is a manner that was only 60% inhibitable by soluble knob protein (CSH poster 184), whereas viruses expressing the wild type protein are about 90% inhibitable. This was given as evidence that the interaction of GRP with its receptor is supporting infection of the host cells.

In one illustrative embodiment, to improve this adenovirus system using the methods of the invention, a mutant fiber protein or a domain replacing the knob that has lost the ability to bind its native receptor is generated. Generation of evolved fiber sequences by recursive recombination generates a new adenovirus fiber or knob-associated ligand with a new specificity. Alternatively, libraries of mutant sequences can be inserted onto the C-terminus of the knob in a manner analogous to the GRP construct described above. Libraries of potential ligands can be randomly inserted throughout the "staff" and/or "knob" domain. The entire knob can be randomly mutagenized and selected for infection of desired targets. Other exposed viral proteins, such as penton or hexon proteins, can be modified with libraries of insertion mutants. Libraries encoding short protein sequences can be inserted in

to adenovirus hexon protein and expressed on the surface of the adenovirus virion as part of the hexon (Crompton (1994) *J. Gen. Virol.* 75:133-139). Next, these modified viruses comprising the recursively evolved viral proteins are used to infect target cells. Diversity and modifications in viral protein affecting adenovirus tropism are selected for by plaque formation, or by cell sorting (FACS), which can be based on transient expression of a reporter gene such as GFP.

Interaction of the fiber penton protein with an integrin on the target cell surface, the alpha-v-integrin, provides a cell-virus stabilizing interaction (it is known that one cannot totally inhibit adenovirus infection with soluble fiber knob protein). In the absence of fiber penton-cell integrin interaction, there is a lower level of viral infectivity. As a result of this complexity in the mechnism which determines the cell specificity of adenovirus, the methods of the invention are used to coevolve multiple genes or domains on adenovirus which interact with their cognate receptors on target cells, such as the penton fiber domain which interacts with target cell alpha-v-integrin. Consequently, recursive sequence recombination of chosen viral genes, or of the whole virus, is a particularly useful tool with which to rapidly evolve tissue-specific adenovirus.

In another illustrative embodiment, the highly developed M13 technology is used to evolve peptide ligands for receptors of interest on target cells. Standard phage display library technology is used to screen for peptide ligands capable of binding purified receptor. Alternatively, the libraries can be screened by panning against cells. The affinity of these ligands is rapidly evolved in M13. Pools of evolved ligands are then engrafted onto target sites on adenovirus, for example, C-terminal fusions to fiber protein. This couples the power of M13 selection to the adenovirus system, making it possible to make libraries of the size that could not be made with M13 alone.

Screening is accomplished by contacting viruses containing recombinant segments with a first population of cells for which infection by the virus is desired and a second population of cells, for which infection is not desired. Viral genomes recovered from the first population of cells are enriched for recombinant segments conferring specificity for that cell type. The first and second populations of cells can be present in different tissues in an organism. For example, one can infect a whole organism with the virus and recover recombinant segments from a subset of blood cells (this being the cell type for which infection is desired). Alternatively, one can infect a whole organism, including humans,

suffering from a natural or induced cancer with virus and recover recombinant segments from the cancer cells. In a further variation, the first and second population of cells are co-cultivated with the virus in mixed cell culture. The two cell types, if they are not readily distinguishable by microscopic examination, can be distinguished by expression of a marker, such as green fluorescent protein or cell surface receptor in one cell type. In the initial round of screening, the existing host cells are usually present in excess (*e.g.*, a ratio of 90% existing host cells to 10% desired target cells). The proportion of desired target cells can be increased in successive rounds.

The recombinant segments recovered from the population of cells for which infection is desired are used as substrates in the next round of recombination. Subsequent rounds of screening are performed by the same principles.

In a variation of the above approach, a eukaryotic or bacterial virus is modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, the EGF family of proteins encompasses several polypeptides such as epidermal growth factor (EGF), transforming growth factor alpha (TGF alpha), amphiregulin (AR) and heregulin (HRG-beta 1), which regulate proliferation in breast cancer cells through interaction with membrane receptors. Han (1995) *Proc. Natl. Acad. Sci. USA* 92:9747-9751, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor.

This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Binding specificity of ligand to receptor can be optimized by recursive recombination of the segment of the viral genome encoding the ligand, and screening using first and second populations of cells as discussed above.

Although viral vectors are most amenable to evolution/recursive recombination to acquire new or altered tissue specificity, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells. Alternatively, variants of nonviral vectors can

be recombined without prior knowledge of sequences that might mediate uptake. For example, the starting substrates can be random sequences. Recombination products are contacted with first and second populations of cells as described above under similar conditions to those contemplated for use of the vector. For example, if a vector is to be used packaged in liposomes, screening is performed with vectors containing recombinant segments packaged as liposomes. Again, vectors containing recombinant segments are recovered from the population of target cells and these segments are used in the next round of recombination.

(I) Improved Uptake of DNA Mediated by Evolved DNA Binding Proteins

The efficiency and specificity of uptake of vector nucleic acid uptake by a given cell type can be improved by coating the vector with an evolved/recursively recombined and modified protein that binds to the nucleic acid. The vector can be contacted with the modified protein *in vitro* or *in vivo*. In the latter situation, the protein is expressed in cells containing the vector, optionally from a coding sequence within the vector. The nucleic acid binding proteins to be evolved usually have nucleic acid binding activity but do not necessarily have any known capacity to enhance or alter nucleic acid DNA uptake.

In this embodiment, DNA binding proteins that are modified by the methods of the invention include transcriptional regulators, enzymes involved in DNA replication (e.g., recA) and recombination, and proteins that serve structural functions on DNA (*e.g.*, histones, protamines). Other DNA binding proteins can include the phage 434 repressor, the lambda phage cI and cro repressors, the *E. coli* CAP protein, myc, proteins with leucine zippers and DNA binding basic domains such as fos and jun; proteins with 'POU' domains such as the Drosophila paired protein; proteins with domains whose structures depend on metal ion chelation such as Cys_2His_2 zinc fingers found in TFIIIA, $Zn_2(Cys)_6$ clusters such as those found in yeast *Gal4*, the Cys_3His box found in retroviral nucleocapsid proteins, and the $Zn_2(Cys)_8$ clusters found in nuclear hormone receptor-type proteins; the phage P22 Arc and Mnt repressors (*see* Knight (1989) *J. Biol. Chem.* 264:3639-3642; Bowie (1989) *J. Biol. Chem.* 264:7596-7602). RNA binding proteins are reviewed by Burd (1994) *Science* 265:615-621, and include HIV Tat and Rev.

As in other embodiment of the invention, evolution of DNA binding proteins toward acquisition of improved or altered uptake efficiency is effected by recursive cycles of recombination and screening. The starting substrates can be nucleic acid segments encoding natural or induced variants of one or nucleic acid binding proteins, such as those mentioned

above. The nucleic acid segments can be present in vectors or in isolated form for the recombination step. Recombination can proceed through any of the formats described in Section II.

For screening purposes, the recombined nucleic acid segments should be inserted into a vector, if not already present in such a vector during the recombination step. The vector encodes a selective marker capable of being expressed in the cell type for which uptake is desired. If the DNA binding protein being evolved recognizes a specific binding site (e.g., *lac*I binding protein recognizes *lac*O), this binding site can be included in the vector. Optionally, the vector can contain multiple binding sites in tandem.

The vectors containing different recombinant segments are transformed into host cells, usually *E. coli*, to allow recombinant proteins to be expressed and bind to the vector encoding their genetic material. Most cells take up only a single vector and so transformation results in a population of cells, most of which contain a single species of vector. After an appropriate period to allow for expression and binding, cells are lysed under mild conditions that do not disrupt binding of vectors to DNA binding proteins. For example, a lysis buffer of 35 mM HEPES {pH 7.5 with KOH}, 0.1 mM EDTA, 100 mM Na glutamate, 5% glycerol, 0.3 mg/ml BSA, 1 mM DTT, and 0.1 mM pMSF) plus lysozyme (0.3 ml at 10 mg/ml) is suitable (see Schatz et al., US 5,338,665). The complexes of vector and nucleic acid binding protein are then contacted with cells of the type for which improved or altered uptake is desired under conditions favoring uptake (e.g., for eukaryotic cells, recipient cells can be treated with calcium phosphate or subjected to electroporation). Suitable recipient cells include the human cell types that are common targets in gene therapy, discussed elsewhere in this application.

After incubation, cells are plated with selection for expression of the selective marker present in the vector containing the recombinant segments. Cells expressing the marker are recovered. These cells are enriched for recombinant segments encoding nucleic acid binding proteins that enhance uptake of vectors encoding the respective recombinant segments. The recombinant segments from cells expressing the marker can then be subjected to a further round of selection. Usually, the recombinant segments are first recovered from cells, *e.g.*, by PCR amplification. The recombinant segments can then be recombined with each other or with other sources of DNA binding protein variants to generate further

recombinant segments. The further recombinant segments are screened in the same manner as before.

In a variation of the above procedure, a binding site recognized by a DNA binding protein can be evolved instead of, or as well as, the DNA binding protein. DNA binding sites are evolved by an analogous procedure to DNA binding proteins except that the starting substrates contain variant binding sites and recombinant forms of these sites are screened as a component of a vector that also encodes a DNA binding protein.

Evolved nucleic acid segments encoding DNA binding proteins and/or evolved DNA binding sites can be included in gene therapy vectors. If the affinity of the DNA binding protein is specific to a known DNA binding site, it is sufficient to include that binding site and the sequence encoding the DNA binding protein in the gene therapy vector together with such other coding and regulatory sequences are required to effect gene therapy. In some instances, the evolved DNA binding protein may not have a high degree of sequence specificity and it may be unknown precisely which sites on the vector used in screening are bound by the protein. In these circumstances, the gene therapy vector should include all or most of the screening vector sequences together with additional sequences required to effect gene therapy.

An exemplary selection scheme is shown in Figure 2. The lower left portion of the Figure shows two vectors, each having the same marker and DNA binding site, the vectors differing in the recombinant segment encoding a DNA binding protein. The vectors are transfected into *E. coli* cells. The vectors are expressed in the cells to produce DNA binding proteins, which differ between the different cells. The recombinant binding proteins complex with the vectors encoding them and these complexes are preserved after cell lysis. The complexes are then contacted with a recipient eukaryotic cell. The eukaryotic cell bears several different cell surface receptors, one of which can interact with one of the DNA binding proteins to facilitate uptake of DNA. Selection for expression of the selection marker on the vector identifies cells transformed with vector. These cells are enriched for recombinant segments conferring enhanced DNA uptake.

(J) Improved Intracellular Stability of a Vector

Vectors with greater and improved cell retention, intracellular stability and expression properties can be developed using the recursive recombination methods of the invention. In many gene therapy methods, it is desirable that the vector be stably maintained

in target cells and thereby be capable of indefinite expression. This is the case for both viral and nonviral vectors. Substrates and recombination formats for evolution of vectors toward improved retention can be chosen according to the principles described above. If the substrates are fragments of vector genomes, the recombination products are reinserted into vector genomes before screening. The vector genomes can often contain a selective marker replacing or fused to the therapeutic coding sequence carried by the vector in actual use. For screening, vector genomes containing recombinant segments are introduced into cells, if they are not already so present as a result of *in vivo* recombination. The cells are grown for a number of generations without selection for the marker, thereby reflecting the situation *in vivo*, in which it is typically not possible to select for retention of a therapeutic gene. After an appropriate period of growth, selection for the marker is applied and surviving cells recovered. These cells can contain vectors harboring recombinant segments conferring the property of improved retention (*i.e.*, recombinant segments stably maintained) in a cell. In some instances, the properties of improved retention, at least in part, a consequence of improved, more stable integration into the cellular genome. Recombinant segments having the property of improved replication, retention and/or stability are recovered from cells, and subjected to a further round of recombination, either with each other and/or with fresh substrates to generate further recombinant segments. These are screened in the same manner as the previous recombinant segments.

(K) Reduced Immunogenicity of Vectors

Protein and nucleic acid sequences with reduced immunogenicity can be developed using the recursive recombination methods of the invention. Immunogenicity is a particular concern with viral vectors, since a host immune response, including CTL mediated and humoral responses, can prevent a virus from reaching its intended target particularly in repeated administrations. Cellular immune responses preventing a virus from reaching its intended target can also be induced against nonviral vectors administered in naked form or shielded with a coat such as liposomes.

Host immune responses which eliminate infected cells is also a major problem in gene therapy. CTLs are primarily responsible for the elimination of infected cells, although the problem can also be partly or entirely antibody-mediated. The recursive recombination methods of the invention can be used to modify a virus to reduce this (primarily cellular) immunity against virally infected cells. In a variation of this embodiment,

for adenovirus-mediated gene transfer, adenovirus late gene expression is reduced by mutations induced by the methods of the invention to reduce CTL responses which contribute to the elimination of virus-infected cells. Thus, the problem of transient retention of virus which can be seen in adenovirus-mediated gene transfer is alleviated.

Substrates and formats for recombination generally follow the principles discussed above. In general, regions of the viral genome encoding outer surface proteins provide the most likely initial substrates for evolution toward reduced immunogenicity. Alternatively, the whole vector genome can be included as an initial substrate for recombination. Recombinant viral genomes should be packaged as viruses before screening, and nonviral genomes should be prepared in the proposed composition for therapeutic administration (*e.g.*, liposomes).

Viruses containing recombinant genomes or nonviral genomes appropriately formulated are administered to a mammal, such as a mouse, rat, rabbit, pig, horse, primate or human, and surviving viruses or nonviral genomes are recovered after an appropriate interval. Often the administration is i.v. and surviving viruses and nonviral genomes are recovered from the blood. Surviving viruses and nonviral genomes are enriched for recombinant segments conferring the property of reduced immunogenicity. These recombinant segments are used as some or all of the substrates in the next round of recombination. Subsequent rounds of selection follow the same format.

In a variation of the above format, antibodies are collected from mammals immunized with the viral library, and immobilized on a column. Another aliquot of the viral library, or a derivative library resulting from a further round of recombination, can then be applied to the column and viruses passing through the column collected. These viruses are enriched for viruses with low immunogenicity.

In a variation of this method for nonviral vectors, the therapeutic expression product of the vector is expressed as a fusion protein joined to a DNA binding protein that has affinity for a sequence on the vector. In this way, at least some of the expression product is maintained in physical proximity with the vector producing it. Thus, immune responses directed against the expression product also remove the vector sequence. Accordingly, recovery of vector sequences surviving a period of time in an animal, enriches both vector sequences that themselves have low immunogenicity and which encode expression products with low immunogenicity.

(L) Reduced Toxicity of Vectors

Protein and nucleic acid sequences with reduced cellular toxicity can be developed using ther recursive recombination methods of the invention. Toxicity caused by viral gene expression is sometimes a concern when using viral vectors in gene therapy. The methods of the invention can be used to induced and select for multiple combinations of mutations blocking viral DNA replication and gene expression *in vivo*. To produced the crippled viruses in vitro, these mutations should be conditional mutations, such as temperature sensitive or nonsense mutations so that the mutant viruses can be propagated in vitro under permissive conditions. The multiplicity and hence redundancy of the conditional mutations prevents the mutant virus from reverting back to the wildtype genotype or phenotype.

(M) Improved Specificity of Integration

Vector sequences with improved specificity of integration can be developed using the recursive recombination methods of the invention. For example, AAV is known to integrate preferentially at a site in chromosome 19q13.3. Integration at this site is advantageous since the presence of an exogenous DNA sequence at this site does not appear to have any adverse effect on expression of endogenous cellular genes. It is therefore desirable to be able to increase the specificity of AAV to this site.

The starting substrates for recombination are AAV vectors including at least ITRs and, optionally, a *rep* gene, since the latter may have a role in site-specific recombination. Genes from other viruses known or believed to have a role in site specific integration can also be included. Preferably the genomes include a marker sequence. Recombination proceeds through any of the recombination formats previously discussed to produce a library of AAV viruses having different recombinant segments in their genomes. The AAV viruses are used to infect appropriate target cells. Cells having taken up AAV DNA can be recognized from expression of the marker. Genomic DNA is isolated from these cells, and a region centered on the intended site of integration is amplified by PCR. The amplified regions are enriched for recombinant segments conferring the desired property of site-specific integration. These recombinant segments form the starting materials for the next round of recombination.

Analogous principles apply to other viral vectors and, indeed, nonviral sequences and vectors. For example, as discussed above, one embodiment of the invention

utilizes site-specific integration systems to target a recombinant sequence of interest to a specific, constant location in the genome. A preferred embodiment uses the Cre/LoxP or the related FLP/FRT site-specific integration system. The Cre/LoxP system uses a Cre recombinase enzyme to mediate site-specific insertion and excision of viral or phage vectors into a specific palindromic 34 base pair sequence ("LoxP site"). The recursive sequence recombination methods of the invention can be used to modify these systems, such as to improve specificity of integration, create alternate, specific sites of integration, modify recombinase activity, and the like.

In a further embodiment, it is not necessary that the starting vector have any preferred integration site. If this is the case, a suitable chromosomal site unlikely to interfere with expression of other genes is chosen, and successive cycles of recombination and selection performed until a vector has evolved to integrate preferentially at that site.

(N) Improved Resistance to Microorganisms

The recursive recombination methods of the invention can also be used to develop new or improve upon known inhibitors of microbial and viral infection, including trans-dominant inhibitors of microbial and viral replication and gene expression. In some gene therapy applications, the vector can encode a product that is an inhibitor to a microorganism, such as a virus. Because of the complexity of viral life cycles and the intrinsic mutability of viruses, recursive sequence recombination is a practical tool for evolving protective antiviral constructs with improved potency and/or new or improved specificities. This can be accomplished using any variety of mechanisms. For example, the gene therapy vector can encode an antisense RNA that blocks expression of a viral or other pathogen's mRNA. The antisense RNA can be designed to bind to a key regulatory sequence, such as a promoter, or to the coding sequence, or both. Alternatively, the vector can encode a protein that is inhibitory to the replication or gene expression of a pathogen. For example, a number of gene therapy strategies have been designed with the intent of inhibiting HIV-1 replication in mature T cells. As T cells are products of hematolymphoid differentiation, insertion of antiviral genes into hematopoietic stem cells serves as a vehicle to confer long-term protection in progeny T cells derived from transduced stem cells. One such "cellular immunization" strategy utilizes the gene coding for the HIV-1 rev trans-dominant mutant protein RevM10 which has been demonstrated to inhibit HIV-1 replication in T-cell lines and in primary T cells; as described in Bonyhadi (1997) *J. Virol.* 71:4707-4716; Nabel

(1996) *Gene Therapy*, abstract 361, CSH. HIV-1 tat and rev mutants have also been suggested as potential intracellular, trans-dominant inhibitors of HIV-1 replication, Caputo (1997) *Gene Ther.* 4:288-295. Another candidate for development by the methods of the invention is the trans-acting transcriptional regulatory protein I kappa B alpha, which can act as a cellular inhibitor of human retroviral replication through a mechanism independent of its effect on HIV transcription, see Wu (1995) *Proc. Natl. Acad. Sci. USA* 92:1480-1484. Repeats of inhibitors derived from viral fragments, such as poly-TAR constructs, can also be used as inhibitors of HIV-1 gene expression. TAR is an RNA stem-loop structure bound by activators or inhibitors of HIV-1 gene expression. TAR can be used to mediate (for example, saturate) cellular factor/RNA interactions, and it has been suggested that transcriptional activators (such as Tat) action might be inhibited by such competing TAR reactions *in vivo*; see Baker (1994) *Nucleic Acids Res.* 22:3365-3372. The recursive recombination methods of the invention can develop and improve upon these and related intracellular inhibitory systems

There are also many examples where a protein from one virus or viral product can be inhibitory to the development of another. Woffendin (1994) *Proc. Natl. Acad Sci. USA* 91:11581-11585. In particular, at least one protein from adeno-associated virus (AAV) is known to be inhibitory to HIV. The large rep gene products, Rep78 and Rep68, of AAV are pleiotropic effector proteins which are required for AAV DNA replication and the trans-regulation of AAV gene expression. Apart from these essential functions, these rep products are able to inhibit the replication and gene expression of HIV-1 and a number of DNA viruses. Batchu (1995) *FEBS Lett.* 367:267-271; Antoni (1991) *J. Virol.* 65:396-404. The recursive recombination methods of the invention can develop new and improve upon these inter-viral inhibitory proteins.

The present invention provides a means for improving the inhibitory qualities of the anti-sense RNAs and proteins described above and also for identifying new inhibitory agents. The improvement to known inhibitory reagents can reside in several aspects, such as improved expression, improved stability or altered fine-binding specificity. It is not necessary in the present methods to know which of these contributory properties is being improved; rather the selection is for the ultimately desired property of microorganism resistance.

For evolution of known inhibitory agents, substrates for recombination and recombination formats are selected according to the principles discussed above. The

substrates can be viral vector genomes or the parts thereof encoding the inhibitory agents and associated regulatory sequences. Initial diversity in recombination substrates can be natural or induced. After a round of recombination, the recombinant segments are introduced into cells (if they are not already in cells as a result of *in vivo* recombination) and the cell are contacted with the microorganism for which protection is desired. Cells surviving exposure to the microorganism are enriched for recombinant segments conferring resistance to the microorganism. These recombinant segments form some or all of the substrates for the next round of recombination.

Similar principles can be applied for *de novo* identification of inhibitory agents to be expressed from gene therapy vectors. More rounds of recombination and screening can be required to obtain satisfactory results. For example, sequences coding for viral proteins from the virus to be inhibited or other viruses provide suitable initial substrates for recombination. The coding sequences can be obtained from the same or different viruses and natural diversity can be augmented by inducing additional mutations, *e.g.*, by error-prone PCR, as described above. Recombination and screening are also performed as described above.

In an illustrative embodiment, a library of mutants is constructed based on candidate construct(s), examples of which are described above. The libraries are transduced or transfected into target cells. The cells are challenged with the microorganism of interest. Resistant cells are isolated based on, for example, survival against cytopathic virus or lack of expression of viral encoded genes, which can include inserted marker genes such as GFP. These methods are used to detect cells in which viral replication or gene expression has been blocked. FACS or panning with an antibody against a virally encoded or induced surface epitope is used in a positive selective step. Genes encoding resistance factor are recovered, for example, by PCR. The recovered genes can be subjected to further rounds of recursive sequence recombination, as described above, until a desired level of protection against the microorganism is achieved

Further illustrative examples of anti-viral mechanisms which can be improved by the methods of the invention include anti-viral ribozyme systems. For example, one or more ribozymes can be targeted against a viral RNA. Adenoviruses have been used to deliver anti-hepatitis C ribozymes; see Lieber (1996) *J. Virol.* 70:8782-8791; Ohkawa (1997) *J. Hepatol.* 27:78-84. HIV-1 Rev response element (RRE) region-specific hammerhead

ribozymes will completely inhibit HIV-1 replication, see Duan (1997) *Gene Ther* 4:533-543. Sendai virus polycistronic P/C mRNA can also be cleaved by ribozymes; Gavin (1997) *J. Biol. Chem.* 272:1461-1472.

Anti-viral cytokines can also be improved by the methods of the invention. For example, wild type or chimeras of wild type interferons such as the IFN alpha 17, IFN beta and IFN gamma constructs can be subjected to recursive sequence recombination. These sequences can placed be under the control of a virus-activated promoter, such as an HIV mini-LTR; see Mehtali (1996) *Gene Therapy*, abstract #364, CSH. For example, cell lines stably carrying IFN transgenes under the positive control of the HIV-1 Tat protein are highly resistant to HIV-1 replication *in vitro*. This antiviral resistance is associated with a strong induction of IFN synthesis immediately following the viral infection. However, IFN-gamma-transfected cells permitted HIV-1 infection *in vivo* despite the induction of a high level of IFN-gamma secretion, see Sanhadji (1997) *AIDS* 11:977-986. The methods of the invention can be used to develop this anti-viral system for potency and effectiveness *in vivo*.

The methods of the invention can be used to develop single chain or Fab antibody fragments directed intracellularly to viral components; Marasco (1996) *Gene Therapy*, abstract 160, CSH. For example, one strategy for somatic gene therapy to treat HIV-1 infection is by intracellular expression of an anti-HIV-1 Rev single chain variable fragment (Sfv); Duan (1997) *Gene Ther, supra*. Intracellular expression of Sfvs which bind to HIV integrase catalytic and carboxy-terminal domains results in resistance to productive HIV-1 infection. This inhibition of HIV-1 replication is observed with Sfvs localized in either the cytoplasmic or nuclear compartment of the cell. See Levy-Mintz (1996) *J. Virol.* 70:8821-8832. The expression of anti-reverse transcriptase (RT) Sfv in T-lymphocytic cells specifically neutralizes the RT activity in the preintegration stage and affects the reverse transcription process, an early event of the HIV-1 life cycle. Blocking the virus at these early stages dramatically decreased HIV-1 propagation, as well as the HIV-1-induced cytopathic effects in susceptible human T lymphocytes, by impeding the formation of the proviral DNA. See Shaheen (1996) *J. Virol.* 70:3392-3400. The methods of the invention can further develop the potency and range of such anti-viral, intracellular antibody fragments.

Improved virus-binding aptamers or peptide ligands directed to viral components, as those described above, can also be further developed by the methods of the

invention. For example, RNA aptamers that recognize a peptide fragment of human HIV-1 Rev were found to bind the free peptide more tightly than a natural RNA ligand, the Rev-binding element, see Xu (1996) *Proc. Natl. Acad. Sci. USA* 93:7475-7480; Symensma (1996) *J. Virol.* 70:179-187. Aptamer sequences isolated from single-stranded DNA preparations have thrombin inhibitory activity, indicating that thrombin-inhibitory aptamers are present in the mammalian genome and may constitute an endogenous antithrombin system. Analogously, the recursive sequence methods of the invention can be used to further identify, develop and improve aptamer sequences useful as anti-microbial agents, or for gene therapy in general.

(O) Viral Packaging Cell Lines

The recursive sequence recombination methods of the invention can also be used to develop new and improved viral packaging cell lines Viral vectors used in gene therapy are usually packaged into viral particles by a packaging cell line. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. AAV recombinants are generally produced by transient co-transfection methods since it has proven difficult to generate stable packaging cell lines (Maxwell (1997) *J. Virol. Methods* 63:129-136).

The goals in improving packaging cell lines include generating stable packaging cell lines; increasing the yield of AAV vector packaged; decreasing the ratio of AAV progeny to helper virus; and reducing the toxicity of the *rep* gene to the packaging cell, which in turn leads to a greater yield of AAV. The leading candidate genes for evolution/ modification by the methods of the invention are the AAV replication (*rep*) and capsid (*cap*)

genes, which can be present on the AAV helper plasmid. Overexpression of the *rep* gene can decrease AAV DNA replication and severely inhibit *cap* gene expression and reduced *rep* level enhances *cap* gene expression and supports normal rAAV DNA replication. Thus, recursive recombination modification of *rep* genes and their expression can generate increased AAV vector production, see Li (1997) J. Virol. 71:5236-5243.

These and related sequences can be subject to recursive sequence recombination according to the general principles discussed. That is, variant forms of these genes are recombined, either *in vivo* or *in vitro*, and cells containing recombinant segments resulting from recombination are screened for a desired property, such as stable packaging cell lines; yield of packaged AAV; increased viability of cells; or, low yield of helper virus relative to packaged AAV. The same principles can be applied to evolve genes in the helper adenovirus, either concurrently or consecutively with the evolution of AAV genes on the helper plasmid.

Cellular genes in the packaging cell line affecting packaging can also be evolved even without knowing what these genes are. This is achieved by transforming the packaging cell line with a library of genes, some of which will undergo recombination with cognate genes in the packaging cell line. The library of genes can be obtained from another type or species of cell or can be a mixture of several types and species and/or can have diversity induced by processes such as error-prone PCR. Cells containing recombinant genes are screened for improved packaging properties, such as increased yield of AAV virus. Optionally, a further library can be transformed into the cells surviving screening in a previous round. Alternatively, the pool of surviving cells can be divided in two, and DNA isolated from one half and used to transform the other half. In this way, the best recombinant segments identified in the first round of screening undergo recombination with each other in the second round of recombination.

EXAMPLES

Example 1: M13 scFv Library

This example shows *in vivo* panning of libraries of bacteriophage displaying scFv for localization to a predetermined cell type, such as a xenogeneic neoplasm. A scFv antibody-phage display library was constructed as described in Crameri (1996) *Nature Medicine* 2:100-102. After growth of the phage library on *E. coli* TG1 in LB containing 50

μg/ml kanamycin, bacterial cells were removed by centrifugation and the phage precipitated by addition of PEG to 4% and NaCl to 0.5 M final concentration. After one hour incubation on ice, the solution was centrifuged at 8,000 x g for 30 minutes, and the pellet resuspended in Dulbecco's phosphate-buffered saline (DPBS).

Male Sprague-Dawley rats were anesthetized and phage were injected intravenously and blood sampled arterially via ipsilateral femoral arterial catheters. EDTA was used in blood samples to reduce coagulation. Blood samples were taken immediately before administration of phage and at 5, 30, 60, 120, and 240 minutes post-injection of 7.6 x 10^{11} colony forming units. Phage titers were determined by dilution of whole blood in DPBS and infection of *E. coli* TG1 to assay colony forming units of M13. Four repetitions of the protocol were performed. It was found that M13 bacteriophage remained stable and infectious (to *E. coli*) with a half-life of six hours in rat blood after *in vivo* injection.

Example 2: Panning of M13 scFv Library for Specific Localization

A scFv antibody-phage display library is administered to mice having transplantable human tumor grafts. After a suitable incubation time, tumor tissue is harvested and phage are eluted from the harvested tissue by homogenization of the tissue sample.

An aliquot of the recovered phage is subjected to at least one additional cycle of administration and selection *in vivo* by the same protocol.

An aliquot of the recovered phage is used to purify DNA and the recovered DNA is recursivcly recombined by shuffling *in vitro*, and the resultant population of shuffled M13 genomes is introduced into *E. coli* and packaged; a library of shuffled M13 species is recovered and administered to mice for at least one additional cycle of administration and selection *in vivo* by the same protocol.

An aliquot of the recovered phage is used to infect *E. coli* at a high multiplicity of infection to recursively recombine M13 genomes *in vivo* by shuffling, and the resultant population of shuffled M13 genomes is introduced into *E. coli* and packaged; a library of shuffled M13 species is recovered and administered to mice for at least one additional cycle of administration and selection *in vivo* by the same protocol.

Example 3: Evolution of the MGMT gene

This example illustrates evolution of the MGMT gene to confer improved properties for protection of human bone marrow against alkylating agents. The wild-type human MGMT cDNA on a high copy number plasmid was amplified by PCR and randomly fragmented with DNase. Small fragments (50-100bp) were reassembled into full-length fragments by Taq DNA polymerase without outside primers in a process that induces point mutations in a rate proportional to the size of the starting fragments, see Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751. Shuffling the entire gene, which encodes 207 amino acids, allows mutagenesis of all regions of the protein including the functionally important DNA-binding region (Kanugula (1995) *Biochemistry* 34:7113-7119). Full-length fragments were cloned back into the vector and transformed into alkyltransferase-deficient *E. coli* (strain GWR111, *ada ogt*) (Rebeck (1991) *J. Bacteriol.* 173:2068-2076). Relatively large numbers of mutations were created to increase diversity and because inactive variants can be eliminated with stringent genetic selection by alkylating agents. This selection involves treating the bacteria with the methylating agent MNNG three sequential times, each separated by a one-hour recovery period during which the bacteria are allowed to make more MGMT. The triple selection kills cells having inactive MGMT and preferentially selects for proteins having improved expression and/or activity of MGMT.

An improved human MGMT gene was also generated using both natural and unnatural -encoding sequence diversity. Unnatural diversity was created by the random fragmentation of the human *MGMT* (wild-type *MGMT* cDNA was generously provided by Dr. S. Mitra, University of Texas, Galveston; see Tano (1990) "Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine," *Proc. Natl. Acad. Sci. USA* 87:686-690, for cDNA and protein sequences and for residue numbering). This was followed by the reassembly of fragments in a mutagenic DNA shuffling reaction. Active variants, selected for their ability to confer MNNG resistance to alkyltransferase-deficient *E. coli*, were pooled, remutated, and recombined in subsequent cycles of shuffling (the alkyltransferase-deficient (*ada ogt*) *E. coli* strain GWR111 was provided by L. Samson, Harvard University, Cambridge, MA; Rebeck (1991) *J. Bacteriol.* 173:2068-2076). Two cycles of conventional DNA shuffling were used to build up the unnatural diversity.

The wild-type human alkyltransferase (*MGMT*) cDNA was subcloned into pUC118 plasmid (New England Biolabs, Beverly, MA) and a translationally silent *Xho*I site created at coding nucleotide residue number 380 (Tano (1990) *supra*, for residue numbering). The flanking non-coding sequences were removed from that construct and an *E. coli* ribosome-binding site added via PCR amplification with oligos 1 and 2 (see below) and inserted into the *Eco*RI-*Hin*DIII sites of pUC118.

Oligo #1: 5'-GCATCCGAATTCCTTAAGGAGGGGAAAAATGGACAAGGATTG-3'

Oligo #2: 5'-CCGCTAAAGCTTCATACTCAGTTTCGGCCAG -3'

This construct is designated "pFC14." The sequence of the entire *MGMT* gene in pFC14 was verified, as was its ability to complement GWR111. A non-functional dummy vector was constructed by replacing the active site-encoding region between the *Xho*I and *Pin*AI sites (nucleotide residue numbers 380 to 521 (Tano (1990) *supra*, for residue numbering) with a synthetic stuffer duplex made by annealing oligos 3 and 4 (below).

Oligo #3: 5'-TCGAGCCCCAGGCCTCCGCA-3'

Oligo #4: 5'-CCGGTGCGGAGGCCTGGGGC-3'

The inactivity of this gene was verified by its inability to complement GWR111. The dummy vector, with the shortened *MGMT* removed, was used as a cloning vector for library construction to reduce the possibility of contamination by wild-type *MGMT*.

The general procedure for creating randomized gene libraries by random fragmentation and reassembly was used as described in Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; and Stemmer (1994) *Nature* 370:389-391. The starting material was a 1.2 kbp PCR product made from pFC14, generated using the outside primers oligo #5: 5'-AAGAGCGCCCAATACGCAAA-3', and oligo #6: 5'-TAGCGGTCACGCTGCGCGTAA-3', and *Taq* DNA polymerase (Promega). This product contained the human *MGMT* plus pFC14 flanking sequence from which 50-300 bp random fragments were prepared and reassembled with *Taq* DNA polymerase, as in Stemmer (1994) *Proc. Natl. Acad .Sci. USA, supra*, and Stemmer (1994) *Nature, supra*. Reamplification with the nested primers oligo #7: 5'-ATGCAGCTGGCACGACAGGTTT-3' and oligo #8: 5'-TACAGGGCGCGTACTATGGTT-3', gave a 980-bp fragment which was treated with *Eco*RI and *Hin*DIII. The resulting 650-bp fragment was ligated into the dummy vector described above. The ligation mixture was electroporated into GWR111, yielding libraries of ~10^5 per

cycle from which active clones were selected. Selection was done as described in Christians (1996) *Proc. Natl. Acad .Sci. USA* 93:6124-6128, with the exception of omission of the inducer isopropyl-beta-thiogalactopyranoside. Bacteria in culture were treated with 3 sequential doses of MNNG, each separated by a 1-hour recovery period. After the third dose all cells were spread on plates. The next day colonies were pooled, and the *MGMT* DNA for the next cycle was prepared by PCR with oligos #5 and #6 (above). This procedure was repeated for a total of 6 cycles. The MNNG treatment was made progressively more stringent as the shuffling progressed, starting at 3 x 10 ug/ml MNNG up to as much as 50 ug/ml in later cycles. Likewise, fewer colonies were picked for shuffling in later cycles.

The natural diversity of four known mammalian alkyltransferases - rat, mouse, hamster, and rabbit - was also used to generate sequence diversity in the improved human MGMT gene. An alignment of their protein sequences, as shown in Figure 4, reveals regions of extensive homology as well as regions of diversity. There exist 2×10^{28} combinations of known natural amino acid substitutions from mammalian alkyltransferases (52 positions with 2 amino acids represented, 24 positions with 3 amino acids, and 2 positions with 4 amino acids = $2^{52} \times 3^{24} \times 4^{2}$). This diversity was exploited through the use of 21 degenerate oligonucleotides (Figure 3). These oligos were mixed together in equal proportions to create one diverse pool, which was mixed with the DNA fragments during the reassembly reactions in the third and fourth cycles. Several different molar ratios of oligos:fragments were made, and it was observed that high concentrations of oligonucleotides inhibited reassembly, probably because the large number of base pair mismatches overwhelmed the polymerase. Of those mixtures giving proper reassembly, as judged by correct product size after reamplification, the one containing the highest proportion of oligos, a molar ratio of 1 oligo:4 fragments, was chosen for further cycling. Annealing of each oligonucleotide to the human-derived *MGMT* sequence was enabled by 20 nucleotides of homology on both sides flanking the degenerate or non-human sequence. Control PCRs demonstrated that all oligonucleotides were approximately equally capable of hybridizing to the human sequence.

In the third round of shuffling, the oligonucleotides were combined with the sequences generated by oligonucleotides having "unnatural diversity," that is, the pooled human *MGMT* clones that survived cycle 2. Conditions were varied in an attempt to incorporate the oligonucleotides and maximize diversity while maintaining the correct size of

the assembled product. The largest molar ratio of oligonucleotide:fragment to allow correct assembly was 1:4. Because of the limitation in the ratio, the "oligo spiking" was repeated in cycle 4. The pools in cycles 3 and 4 were thus hybrids containing randomly mutated human-derived sequence as well as different combinations of mammalian *MGMT* gene segments. These pools were subjected to selection between cycles. Two final rounds, cycles 5 and 6, of "conventional shuffling," without addition of oligonucleotides, were performed in an attempt to further evolve the hybrid proteins.

Individual clones surviving later cycles were screened for improvement by treating them with a single 40 ug/ml dose of MNNG and comparing survival to untreated samples. The best performing clone, from cycle 4, showed a 10-fold improvement over the wild-type at this dose. Its deduced protein (amino acid) sequence, shown in Figure 5 (SEQ ID NO:2), based on the improved (evolved) nucleotide sequence (SEQ ID NO:1), contains 7 amino acid differences from the wild-type human alkyltransferase (see the seven circled amino acid residues in Figure 5), 5 of which are found in other mammalian alkyltransferases (boxed residues in Figure 4). These 5 amino acid changes presumably were encoded by the oligonucleotides spiked in during cycles 3 and 4. All 5 were encoded by the same degenerate oligonucleotide pool, #7 in Figure 3. The other amino acid changes, Q (gln) to R (arg) at residue number 72 (Q72R) and G (gly) to D (asp) at residue number 173 (G173D) (Tano (1990) *supra*), were not present in the natural diversity and thus were created by the mutagenic shuffling process. In addition, 2 translationally silent nucleotide changes (from the wild type) were detected (see the two underlined nucleic acid residues in Figure 5).

This shuffled mutant was characterized more thoroughly for its activity in *E. coli*. In one set of experiments, cells were treated with graded doses of MNNG and the surviving fraction determined. Plasmids isolated from individual clones surviving the MNNG treatments were retransformed into GWR111. The retransformed clones were screened individually by treating them with a single 40 ug/ml dose of MNNG. The best performing clone was further analyzed three ways: (*i*) The entire *MGMT* DNA sequence was obtained by sequencing the DNA target bidirectionally using fluorescent dye terminator cycle sequencing methods (Applied Biosystems 373A Autosequencer, Foster City, CA); (*ii*) Kill curves were established by treating exponentially growing cells with graded single doses of MNNG in the absence of isopropyl-beta-thiogalactopyranoside and measuring colony-

forming ability relative to untreated controls. Cells harboring the wild-type gene (pFC14) or the V139F mutant (Christians (1996) *Prod. Natl. Acad. Sci. USA, supra*) were treated in parallel for comparison; (*iii*) The alkyltransferase activity of bacterial extracts was quantitated by *in vitro* exposure to calf thymus DNA containing O^6-[^{3}H]methylguanine as described in Bobola (1995) *Molec. Carcinogen* 13:70-80. Some extracts were preincubated with the mammalian alkyltransferase inhibitor O^6-benzylguanine.

Survival was greater than for cells harboring either the wild-type human MGMT or the V139F mutant. The LD_{10}'s, or dose of MNNG giving 10% survival, were: wild-type, 17.5 ug/ml; V139F, 25 ug/ml; and cycle 4 shuffled mutant, 33 ug/ml. In a second set of experiments, bacterial extracts were exposed *in vitro* to an excess of [^{3}H]-methylated DNA substrate, primarily in the form of O^6-methylguanine, to measure total alkyltransferase activity. Average insoluble counts per minute per ug of total protein were: wild-type, 126; V139F, 58; and cycle 4 shuffled mutant, 52. All three proteins were sensitive to the inhibitor O^6-benzylguanine.

Thus, the recursive sequence recombination methods of the invention has successfully generated a new and improved human alkyltransferase protein. The random diversity created by the mutagenic shuffling process was augmented by the diversity provided by nature. Natural diversity was utilized by simply mixing fragments of the human gene with oligonucleotides encoding all of the known mammalian amino acid substitutions. Homology to the human gene in the sequence flanking the regions of diversity facilitated incorporation of the oligonucleotides. The best performing mutant was a hybrid with 7 amino acid differences from the human alkyltransferase, as shown in Figure 5 (SEQ ID NO:1). Two of the mutations arose spontaneously during shuffling, and the other 5 were encoded by the natural diversity, specifically, one of the "spiked oligos" spanning amino acid position 50. Because all oligos were shown by PCR to be capable of hybridizing to the human sequence, it is likely that all were incorporated into the pool at least to some degree.

Previous work with a different system also confirmed that synthetic oligos in such a reaction are incorporated at approximately the expected ratios (Crameri (1996) *Nature Medicine, supra*). Another way to incorporate natural diversity is to isolate or synthesize the cDNA from each of the species and shuffle the entire coding sequences together. This recursive method of breeding natural diversity will improve many related genes from

different organisms as well as gene families within an organism. Furthermore, it can be applied to multiple proteins with related motifs, either structural or functional.

It is difficult to mechanistically rationalize how the amino acid substitutions in the shuffled mutant increase its activity in *E. coli*. None of the amino acid positions mutated in the shuffled mutant was assigned a function in a computer model of the human alkyltransferase based upon the sole alkyltransferase crystal structure, that of the bacterial Ada protein C-terminal fragment (Wibley (1995) *Cancer Drug Design* 10:75-95; Moore (1994) *EMBO J.* 13:1495-1501. The clustering of 5 of the mutations around position 50 is striking, but no known function has been ascribed to this region of the protein. Three of these 5 substitutions are found in all of the other mammalian alkyltransferases. While some substitutions might be neutral, a possibility that can be answered by backcrossing, others might be synergistic, especially those involving charge changes. The proximity of the G (gly) to D (asp) mutation at position number 173 (G173D) (see Tano (1990) *supra*, for residue numbering) to the conserved E (glu) at residue number 172 (E172) might be significant given the proposed involvement in crucial salt-link interactions by E172. An additional acidic residue in the region might enhance this effect.

The power of DNA shuffling is that it is a molecular breeding process that allows for the combination of mutations which incrementally improve many such complex effects without having to model the effects in detail. We have exploited this property to evolve an alkyltransferase that is more potent *in vivo* than the natural enzyme or any reported mutants. This evolved mutant will be very useful in chemoprotection by gene therapy. An improvement over wild-type alkyltransferase is very useful to the clinician by allowing dose escalation of alkylating agents without the corresponding toxicity to the patient. Once-promising alkylating agents which are not used because of severe myelotoxicity might now become clinically acceptable. Even a slight improvement in alkyltransferase *in vivo* is useful, given that positive selection allows a relatively small number of resistant cells to repopulate the bone marrow. This alkyltransferase is further modified to incorporate additional features such as O^6-benzylguanine resistance. The alkyltransferase can also be subjected to additional DNA shuffling and selected for additional improved activity in mammalian cells, such as improved nuclear localization, or better interaction with the eukaryotic chromatin structure.

Example 4: Whole Genome Shuffling of Virus by *In Vivo* Recombination Using Adenovirus-Phagemids

This example demonstrates the construction of an novel adenovirus-phagmid using the recursive recombination methods of the invention which is capable of packaging DNA inserts over 10 kilobases in size. Incorporation of a phage f1 origin using the methods of the invention also generates a novel *in vivo* shuffling format capable of evolving whole genomes of viruses, such as the 36 kb family of human adenoviruses.

The widely used human adenovirus type 5 (Ad5) has a genome size of 36 kb. It is difficult to shuffle this large genome *in vitro* without creating an excessive number of changes which may cause a high percentage of nonviable recombinant variants. To minimize this problem and achieve whole genome shuffling of Ad5, an adenovirus-phagemid was constructed using the methods of the invention.

As outlined in Figure 6, the 36 kb Ad5 genome was divided into two overlapping parts by restriction digestion. Each of the two halves were subcloned into pBR322; the resulting two plasmids designated pAd-R and p-Ad-L. Specifically, an EcoR I ready-made adaptor was first ligated to each end of the linear 36 kb genomic DNA. This ligation product was then digested with BamH I to generate the right half of the Ad5 genome (nucleotide 21,562 to 35,935); and, with EcoR I to generate the left half of the genome (nucleotide 1 to 27,331). The right half 14.3 kb BamH I /EcoR I fragment was then ligated with BamH I /EcoR I digested pBR322 to create Ad-R, and the left half 27.3 kb EcoR I fragment was ligated with EcoR I digested pBR322 to created pAd-L. For gene transfer and safety reasons, the Ad5 E1 region was subsequently deleted from the pAd-L by: first, creating an Afl II restriction site at nucleotide 455 using site directed mutagenesis (changing G residue at position 457 to a T residue, and a C residue at position 459 to an A residue); and, Afl II partial digestion was then performed since there are other Afl II sites in the plasmid. The 24.3 kd Afl II fragment was gel purified, filled in with DNA polymerase I to create blunt ends. It was then ligated with a Swa I linker to simultaneously delete the E1 region between nucleotides 458 and 3533, and insert a unique Swa I site for insertion of foreign genes.

To construct phagemids ssDNA phage f1 replication origin was obtained by PCR from pBluescript II KS(-) phagemid (Stratagene, San Diego, CA) and ligated into the Cla I site of the Ad plasmids (pAd-R and pAd-L-1) by recombinant DNA techniques, as illustrated in Figure 6. The resulting Ad-phagemids were then introduced into a mutator

strain *mut*D5 (see Degnen (1974) *J. Bacteriol.* 117:477-487) to obtain mutations, thus increasing diversity. The spontaneous mutation rate of mutD5 strains is approximately 1.8 x 10^{-6}/base pair/cell/generation (see Fijalkowska (1996) *Proc. Natl. Aca. Sci. USA* 93:2856-2861), which is about 100 fold lower than that of *in vitro* shuffling (see Stemmer (1994)*Proc. Natl. Aca. Sci. USA* 93:2856-2861).

To prepare phagemid phage, these mutated Ad-phagemids were purified from the mutD5 cells and then introduced into a F+ *recA1* strain (XL-1 Blue, Stratagene, San Diego, CA), and the resulting transformants were infected with a helper M13 phage (VCSM13, Stratagene, San Diego, CA) with a multiplicity of infection (MOI) of 10. The *recA1* mutation, which abolishes the recombinase activity of RecA (see Clark (1965) *Proc. Natl. Aca. Sci. USA* 53:451-459), is essential for the stability of the 29 kb pAd-L-f1 during helper phage infection. Stable, high titer (>10^{10} transducing units per ml) stocks of Ad-phagemid phage were obtained. These ssDNA phages carrying the Ad genome were then used to infect a *mut*S 201:Tn5 strain (see Siegel (1982) Mutat. Res. 93:25-33) at high multiplicity to promote recombination *in vivo*. Homologous recombination is particularly efficient between single-stranded forms of intracellular DNA. After replication, the phagemids within the cell behave as regular plasmids and undergo additional plasmid-plasmid recombination during subsequent cell propagation. The shuffled Ad-phagemids were finally recovered and purified from the cells, and used to transfect HeLa cells to generate high titer libraries.

Phagemid vector have been widely used for peptide display, cDNA cloning and site-directed mutagenesis (see Mead (1988) *Biotechnol.* 10:85-102 for review). However, phagemid vector have not been used with large sizes (inserts) of DNA. Conventional phagemid systems have not been used for cloning DNA fragments larger than 10 kilobases or to generate large-sized (>10 kb) ssDNA. The invention's Ad-phagemid has been demonstrated to accept inserts as large as 15 and 24 kilobases and to effectively generate ssDNA of that size. In a further embodiment, larger DNA inserts, as large as 50 to 100 kb are inserted into the Ad-phagemid of the invention; with generation of full length ssDNA corresponding to those large inserts. Generation of such large ssDNA fragments provides a means to evolve, *i.e.* modify by the recursive recombination methods of the invention, entire viral genomes. Thus, this invention provides for the first time a unique phagemid system

capable of cloning large DNA inserts (>10 KB) and generating ssDNA *in vitro* and *in vivo* corresponding to those large inserts.

Example 5. The generation of retroviral vectors carrying mutant drug transporters

A pool of cells expressing a library of variants of ABC transporters is generated by shuffling the wildtype cDNA such as e.g. the MDR1 or cMOAT cDNA as described for the MGMT gene in example 3. The libraries are cloned into a retroviral backbone such as described in PCT/NL96/00195 (filed May 7 1996 published under WO96/41875) followed by transfection into a retroviral packaging cell line. After stable or transient transfection flow cytometric sorting of cells pumping out the drug most efficiently is performed to rapidly select for those cells expressing the desired phenotype from the retroviral construct. If the MDR or cMOAT drug/substrate is fluorescent by itself such as in the case of anthracyclins or rhodamine for MDR1 this can be used to sort cells expressing a desired mutant. In the case of MDR1, fluorescent analogues of AZT (3'-azido--2',3'dideoxythymidine), ddC (2',3'-dideoxycytidine) or etoposide, or BODIPY conjugates of paclitaxel (a taxol equivalent) can be used to viably sort or separate cells negative for these dyes from cells positive for the fluorescent drug and thus negative for a particular MDR1 variant. Optionally flow cytometric sorting is followed by selection for those cells actually resistant to the drug used for flow cytometric sorting or direct cloning by single cell sorting or convential limiting dilution in tissue culture. Because the cells are retroviral packaging cell lines the selected cells can than be tested for the production of retrovirus carrying the mutant version of the ABC transporter under investigation.

An alternative to making recursive recombination libraries from single drug resistance sequences is to subject a complete vector carrying for example MDR1 or cMOAT to recursive recombination. This could be advantageous because the performance of for example a retroviral vector carrying a transgene such as MDR1 is influenced by the transgene polynucleotide sequence itself. Therefore optimal vectors for a given application may be generated by starting from a complete vector including but not limited to the MDR1 retroviral vector disclosed in PCT/NL96/00195.

Example 6: Testing of selected pools of vectors carry mutant drug resistance genes on human hematopoetic stem cells by flow cytometry.

Vectors generated using the methods disclosed herein that carry mutant drug transporter genes are tested for their performance in stem cells by employing flow cytometric assays.

A multi-color flow cytometric assay enables one to study multiple parameters such as differentiation, cycling, amphotropic receptor expression and retroviral vector-mediated transduction concomitantly at a single-cell level using immunophenotyping. The most primitive hematopoietic progenitors to study are the $CD34^{bright}$Lin(CD33, CD38 and CD71)-cells. This candidate hematopoietic stem cell population is identified by staining with monoclonal antibodies, conjugated to two different fluorochromes, and analysed on two emission channels. Another emission channel is used to measure transport activity of mutant drug transporter carried by a recombinant viral vector. Using such a multiparameter flow cytometric analysis drug resistance phenotype of selected MDR1 or cMOAT or MRP1 or MRP3 variants are determined on CD34+lin- cells from human bone marrow or human cord blood cells or human peripheral blood cells. Variants exhibiting significant transport activity in CD34+lin- cells are tested *in vivo* NOD-SCID mice (see example 7).

Example 7: Testing of selected pools of drug resistance genes on human hematopoetic stem cells in vivo

After human patients and non-human primates, the NOD/Scid-Human chimera murine model is the most valid assay to study human primitive hematopoietic cells (Dick et al, Semin Immunol. 8 (4):197-206,1996). By analyzing bone marrow cells from mice transplanted with umbilical cord blood CD34+ cells once a month, high levels of engraftment and multi lineage differentiation are observed as soon as 4 weeks after transplantation Verlinden et al, Blood. 88: 168.,1996). After 6 months, human granulocytes, platelets, lymphocytes and erythrocytes are found in both the murine bone marrow and peripheral blood.

CD34+lin- cells as described under example 6 are isolated using FACS and infected *ex vivo* with vectors generated using the methods disclosed here and that carry mutant drug transporter genes. The infected cells are then infused into irradiated NOD/Scid mice followed by in vivo selection of the transduced cells using the drug by which the mutant

drug resistance gene was isolated. Doing so the *in vivo* performance of the new drug transporter or drug transporter vector or both is assessed by measuring selective outgrowth of the human stem cells as compared to CD34+lin- cells transduced with vector carrying the wildtype drug transporter.

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching. Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention. All patent documents and publications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each item were so individually denoted.

WHAT IS CLAIMED IS:

1. A method of evolving a drug transporter gene, comprising:

(1) recombinating at least first and second forms of the gene differing from each other in at least two nucleotides, to produce a library of recombinant genes;

(2) screening at least one recombinant gene from the library for conferring improved or altered drug resistance;

(3) recombining, as necessary, at least one recombinant gene with a further form of the gene, the same or different from the first and second forms, to produce a further library of recombinant genes;

(4) screening, as appropriate, at least one further recombinant gene from the further library for improved or altered drug resistance;

(5) repeating (3) and (4), as necessary, until the further recombinant gene confers a desired level of improved or altered drug resistance.

2. The method of claim 1, wherein more than one round of screening is performed between successive steps of recombining.

3. The method of claim 1 or 2, wherein the recombinant or further recombinant genes are screened by exposing cells to a drug and selecting surviving cells, the surviving cells being enriched for recombinant or further recombinant genes having the property of conferring improved or altered drug resistance.

4. The method of claim 3, further comprising increasing the concentration of the drug between successive rounds of screening.

5. The method of anyone of claims 1 to 4, wherein the drug is a chemotherapeutic drug.

6. The method of claim 1 or 2, wherein the recombinant or further recombinant genes are screened by detecting efflux from cells of a substrate for a drug transporter encoded by the drug transporter gene or by the recombinant or further recombinant genes and selecting the cells containing low intracellular amounts of said substrate.

7. The method of claim 1 or 2, wherein the recombinant or further recombinant genes are screened by detecting influx into cells of a substrate for a drug transporter encoded by the drug transporter gene or by the recombinant or further recombinant genes and selecting the cells containing high intracellular amounts of said substrate.

8. The method of anyone of claims 3 to 7, wherein the cells are stem cells.

9. The method of anyone of claims 3 to 7, wherein the cells are kidney cells, heart cells, lung cells, liver cells, gastrointestinal or central nervous system cells.

10. The method of any of the aforementioned claims, for use of the recombinant or further recombinant gene in gene therapy.

11. The method of claim 1 or 2, wherein at least one recombining step occurs *in vivo*.

12. The method of claim 1 or 2, wherein at least one recombining step occurs *in vitro*.

Fig. 1

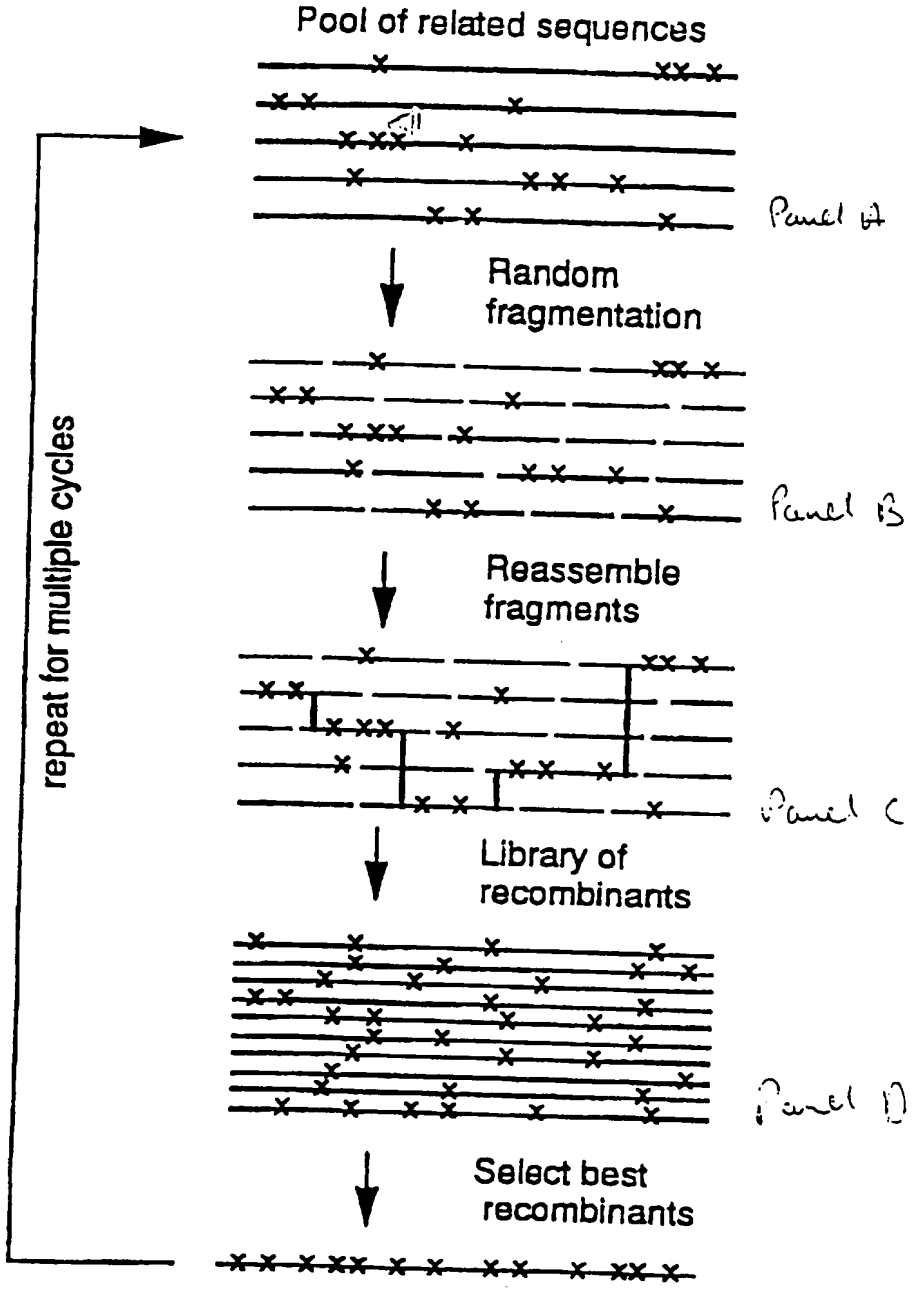

Pool of related sequences
Panel A
Random fragmentation
Panel B
Reassemble fragments
Panel C
Library of recombinants
Panel D
Select best recombinants
repeat for multiple cycles

Fig. 2

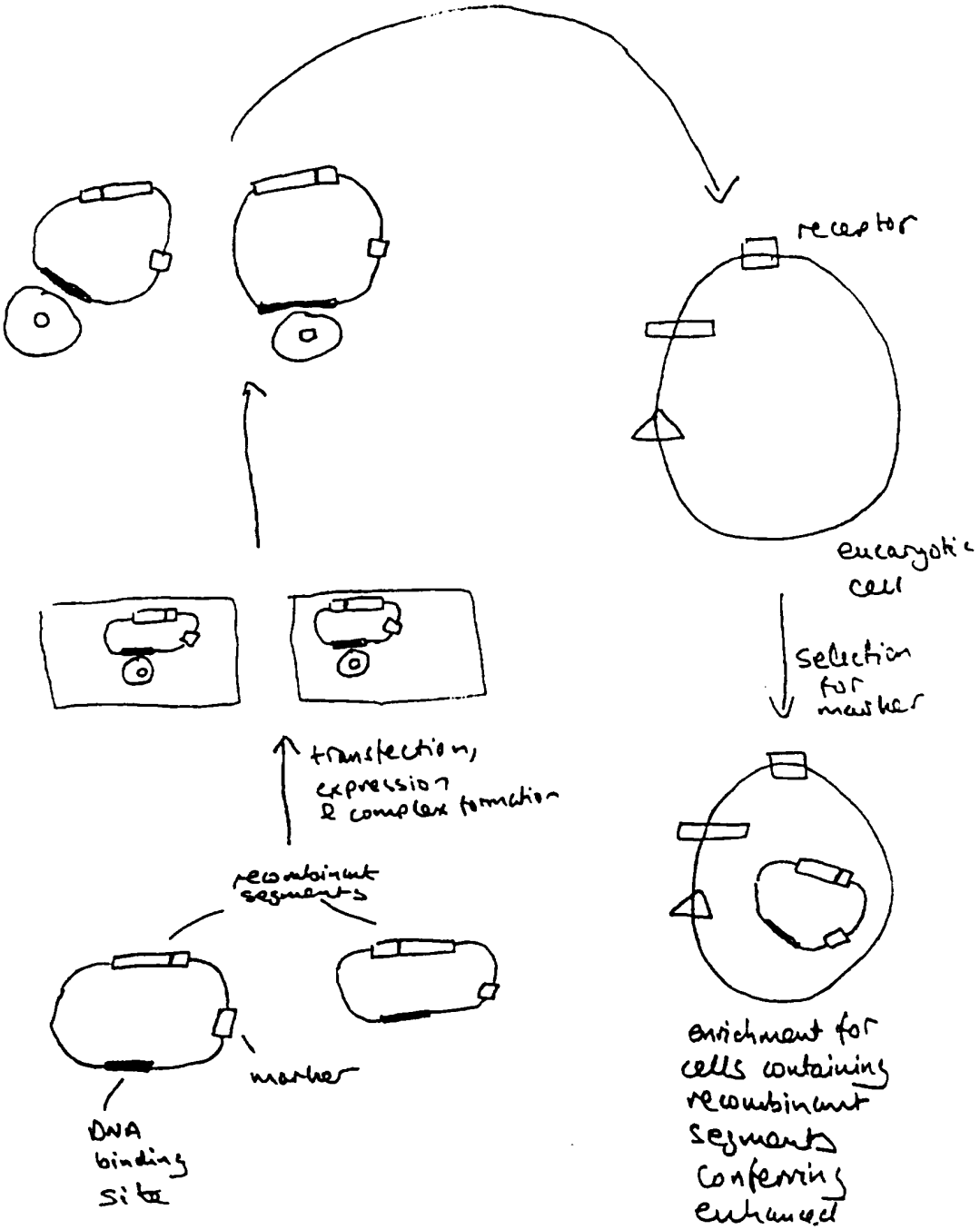

receptor
eucaryotic cell
selection for marker
enrichment for cells containing recombinant segments conferring enhanced DNA uptake
transfection, expression & complex formation
recombinant segments
marker
DNA binding site

Figure 3

Degenerate oligos for spiking mammalian diversity into human MGMT

A1. CCTTAAGGAGGGGAAAAATG GCC GAG AYT TGT AAA ATGAAACGCACCACACTGGA

A2. AAATGGACAAGGATTGTGAA CTG AAA TAC AWK GTG TTC GACAGCCCTTTGGGGAAGCT

A3. AAATGAAACGCACCACACTG SMC AGC CCT TTG GGG GCG ATR GAGCTGTCTGGTTGTGAGCA

A4. TGGAGCTGTCTGGTTGTGAG CGG GGT CTG CAC RGT ATAAAGCTCCTGGGCAAGGG

A5. AGCAGGGTCTGCACGAAATA CGG TTC CTC AGC GGG AAG ACGTCTGCAGCTGATGCCGT

A6. AGCTCCTGGGCAAGGGGACG CCT ARM WCT GAT CCC AMA GAGGTCCCAGCCCCCGCTGC

A7. CTGCAGCTGATGCCGTGGAG GCC CCA GCC WSC CCT GAG KKG CTCGGAGGTCCGGAGCCCCT

A8. CGGTTCTCGGAGGTCCGGAG TCC CTG GTG CAG TGC GAA ACC TGGCTGAATGCCTATTTCCA

A9. TGCAGTGCACAGCCTGGCTG SAW GCC TAT TTC CRA GAG CCCGAGGCTATCGAAGAGTT

A10. ATGCCTATTTCCACCAGCCC KCG GCT ACC CCA GGG CTG CCCGTGCCGGCTCTTCACCA

A11. AGGCTATCGAAGAGTTCCCC TTG CCGGCTCTTCACCATCCCGT

A12. ACCATCCCGTTTTCCAGCAA GAT TCGTTCACCAGACAGGTGTT

A13. AGGTTGTGAAATTCGGAGAA AYG GTT TCTTACCAGCAATTAGCAGC

A14. CAGTGGGAGGAGCAATGAGA ARC AATCCTGTCCCCATCCTCAT

A15. TCATCCCGTGCCACAGAGTG ATC CGC AGC RAC GGA TCC ATT GGCAACTACTCCGGAGGACT

A16. GCAGCAGCGGAGCCGTGGGC CAC TAC TCC GGA GGA CAG GCCGTGAAGGAATGGCTTCT

A17. GGCTTCTGGCCCATGAAGGC TYC CCG AMG AGG CAG CCA GCC TTGGGGAAGCCAGGCTTGGG

A18. GGTTGGGGAAGCCAGGCTTG IST AAG GRC TTA GCT CTG AYT GGGGCCTGGCTCAAGGGAGC

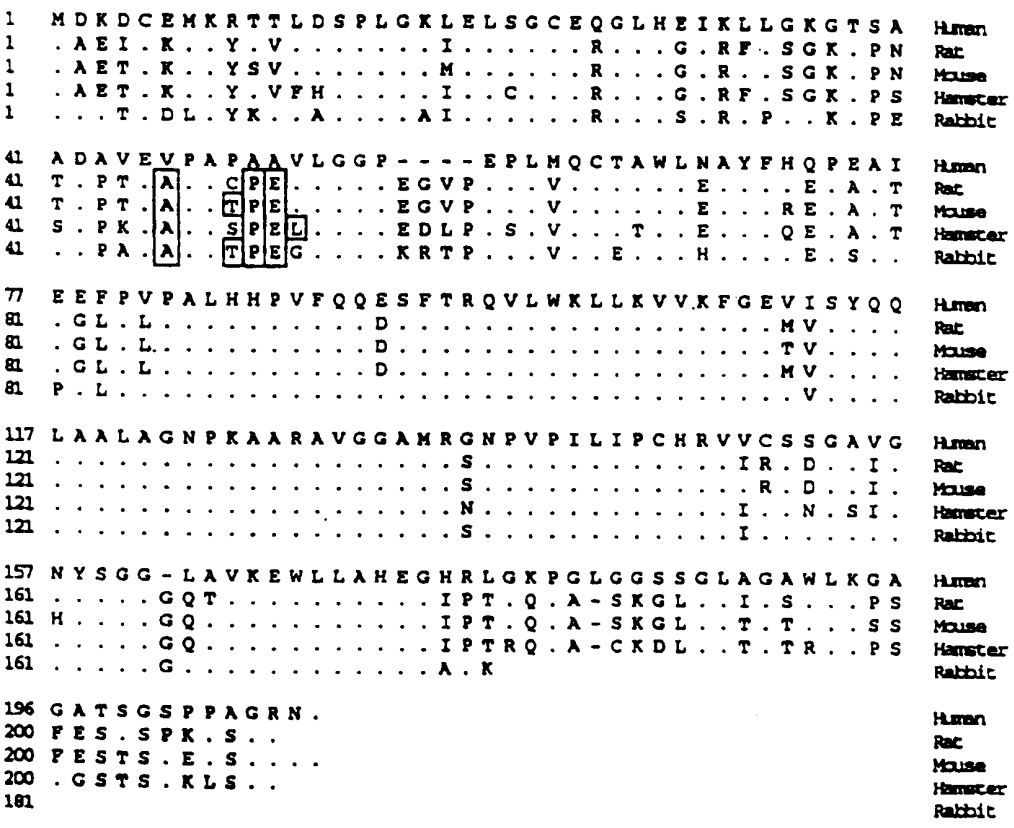
A19. GGAGCTCAGGTCTGGCAGGG WCC CGG CTCAAGGGAGCGGGAGCTAC

A20. TGGCAGGGGCCTGGCTCAAG YCA TCG TTC GRG TCC TCGGGCTCCCCGCCTGCTGG

A21. TCAAGGGAGCGGGAGCTACC ACG AGC CCC RAG CTT TCT GGCCGAAACTGAGTATGAAG

Each oligo contains 20 bp homology to human cDNA 5' and 3', and
up to 21 bp non-homology (underlined) to incorporate all known mammalian diversity.
IUPAC ambiguity codes are used.

Figure 4


1 MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSA Human
1 .AEI.K..Y.V.....I.....R...G.RF.SGK.PN Rat
1 .AET.K..YSV.....M.....R...G.R..SGK.PN Mouse
1 .AET.K..Y.VFH.....I..C...R...G.RF.SGK.PS Hamster
1 ...T.DL.YK..A....AI.....R...S.R.P..K.PE Rabbit
41 ADAVEVPAPAAVLGGP----EPLMQCTAWLNAYFHQPEAI Human
41 T.PT.A..CPE.....EGVP...V.....E....E.A.T Rat
41 T.PT.A..TPE.....EGVP...V.....E...RE.A.T Mouse
41 S.PK.A..SPEL....EDLP.S.V...T..E...QE.A.T Hamster
41 ..PA.A..TPEG....KRTP...V..E...H....E.S.. Rabbit
77 EEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISYQQ Human
81 .GL.L.........D.....................MV.... Rat
81 .GL.L.........D.....................TV.... Mouse
81 .GL.L.........D.....................MV.... Hamster
81 P.L.....................................V.... Rabbit
117 LAALAGNPKAARAVGGAMRGNPVPILIPCHRVVCSSGAVG Human
121 .....................S.............IR.D..I. Rat
121 .....................S.............R.D..I. Mouse
121 .....................N.............I..N.SI. Hamster
121 .....................S.............I..... Rabbit
157 NYSGG-LAVKEWLLAHEGHRLGKPGLGGSSGLAGAWLKGA Human
161 .....GQT.........IPT.Q.A-SKGL..I.S...PS Rat
161 H....GQ.........IPT.Q.A-SKGL..T.T...SS Mouse
161 .....GQ.........IPTRQ.A-CKDL..T.TR..PS Hamster
161 .....G..........A.K Rabbit
196 GATSGSPPAGRN. Human
200 FES.SPK.S.. Rat
200 FESTS.E.S.... Mouse
200 .GSTS.KLS.. Hamster
181 Rabbit

DNA Strider™ 1.2 ### Friday, September 26, 1997 9:56:09 AM

4.6 -> 1-phase Translation

DNA sequence 624 b.p. ATGGACAAGGAT ... GGCCGAAACTGA linear

```
1/1
ATG GAC AAG GAT TGT GAA ATG AAA CGC ACC                                  31/11
M   D   K   D   C   E   M   K   R   T   ACA CTG GAC AGC CCT TTG GGG AAG CTG GAG
61/21                                   T   L   D   S   P   L   G   K   L   E
CTG TCT GGT TGT GAG CAG GGT CTG CAC GAA ATA AAG CTC CTG GGC AAG GGG ACG TCT GCA
L   S   G   C   E   Q   G   L   H   E   I   K   L   L   G   K   G   T   S   A
121/41                                  151/51
GCT GAT GCC GTG GAG GCC CCA GCC ACC CCT GAG CTG CTC GGA GGT CCG GAG CCC CTG ATG
A   D   A   V   E   A   P   A   T   P   E   L   L   G   G   P   E   P   L   M
181/61                                  211/71
CAG TGC ACA GCC TGG CTG AAT GCC TAT TTC CAC CGG CCC GAG GCT ATC GAA GAG TTC CCC
Q   C   T   A   W   L   N   A   Y   F   H   R   P   E   A   I   E   E   F   P
241/81                                  271/91
GTG CCG GCT CTT CAC CAT CCC GTT TTC CAG CAA GAG TCG TTC ACC AGA CAG GTG TTA TGG
V   P   A   L   H   H   P   V   F   Q   Q   E   S   F   T   R   Q   V   L   W
301/101                                 331/111
AAG CTG CTG AAG GTT GTG AAA TTC GGA GAA GTG ATT TCT TAC CAG CAA TTA GCA GCC CTG
K   L   L   K   V   V   K   F   G   E   V   I   S   Y   Q   Q   L   A   A   L
361/121                                 391/131
GCA GGC AAC CCC AAA GCC GCT CGA GCA GTG GGA GGA GCA ATG AGA GGC AAT CCT GTC CCC
A   G   N   P   K   A   A   R   A   V   G   G   A   M   R   G   N   P   V   P
421/141                                 451/151
ATC CTC ATC CCG TGC CAC AGA GTG GTC TGC AGC AGC GGA GCC GTC GGC AAC TAC TCC GGA
I   L   I   P   C   H   R   V   V   C   S   S   G   A   V   G   N   Y   S   G
481/161                                 511/171
GGA CTG GCC GTG AAG GAA TGG CTT CTG GCC CAT GAA GAC CAC CGG TTG GGG AAA CCA GGC
G   L   A   V   K   E   W   L   L   A   H   E   D   H   R   L   G   K   P   G
541/181                                 571/191
TTG GGA GGG AGC TCA GGT CTG GCA GGG GCC TGG CTC AAG GGA GCG GGA GCT ACC TCG GGC
L   G   G   S   S   G   L   A   G   A   W   L   K   G   A   G   A   T   S   G
601/201
TCC CCG CCT GCT GGC CGA AAC TGA
S   P   P   A   G   R   N   *
```

mutant alkyltransferase sequence

Figure 5

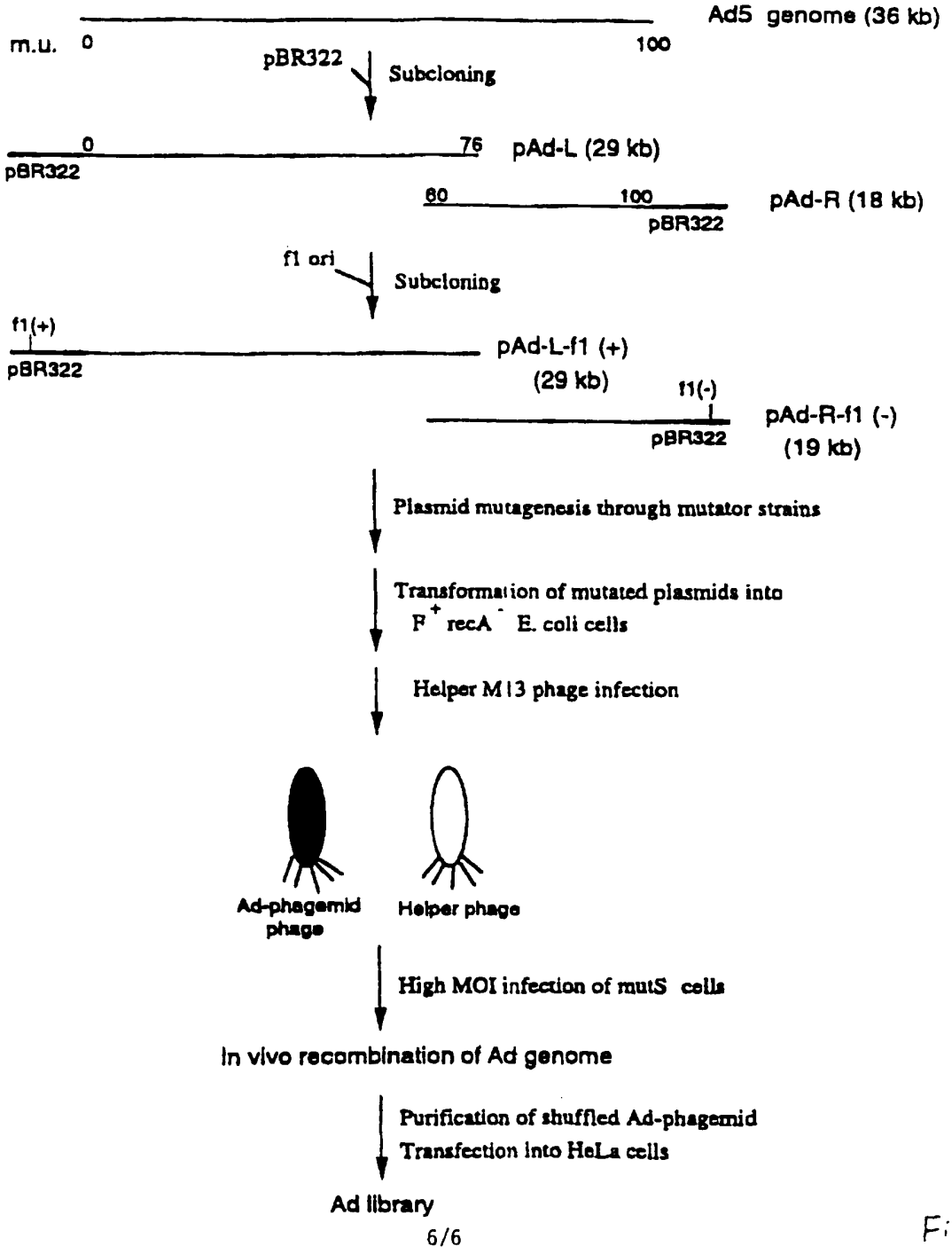

Ad5 genome (36 kb)
m.u. 0
100
pBR322
Subcloning
0
76
pAd-L (29 kb)
pBR322
80
100
pAd-R (18 kb)
pBR322
f1 ori
Subcloning
f1(+)
pBR322
pAd-L-f1 (+)
(29 kb)
f1(-)
pBR322
pAd-R-f1 (-)
(19 kb)
Plasmid mutagenesis through mutator strains
Transformation of mutated plasmids into
F+ recA- E. coli cells
Helper M13 phage infection
Ad-phagemid phage
Helper phage
High MOI infection of mutS cells
In vivo recombination of Ad genome
Purification of shuffled Ad-phagemid
Transfection into HeLa cells
Ad library

Fig 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17302

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C12N 15/00; C12Q 1/68
US CL :435/6, 172.1, 172.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 172.1, 172.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, MEDLINE, BIOSIS, EMBASE, CAPLUS, SCISEARCH
search terms: recombination, exon shuffel?, drug transport?, MDR, library, mutation, mutate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	STEMMER, W. Rapid Evolution of a Protein *in vitro* by DNA Shuffling. Nature. 04 August 1994, Vol. 370, pages 389-391, see entire document.	1-4, 6-9, 11-12 ----- 5, 10
X -- Y	STEMMER, W. Searching Sequence Space Using Recombination to Search More Efficiently and Thoroughly Instead of Making Bigger Cominatorial Libraries. Bio/Technology. June 1995, Vol. 13, pages 549-553, see entire document.	1-4, 6-9, 11-12 ----- 5, 10
Y	STEMMER, W. DNA Shuffling by Random Fragmentation and Reassembly: *in vitro* Recombination for Molecular Evolution. Proc. Natl. Acad. Sci. USA. October 1994, Vol. 91, pages 10747-10751, see entire document.	1-12

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
01 DECEMBER 1997	03 FEB 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer SEAN MCGARRY Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/17302

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BURGER, H. et al. Analysis of MRP Expression in a Large Variety of Human Cancers. Proceedings of the American Association for Cancer Research. March 1995, Vol. 36, page 522, abstract No. 3108, see entire abstract.	1-12
Y	BOLHUIS, H. et al. The Lactococcal *lmrP* Gene Encodes a Proton Motive Force-Dependent Drug Transporter. The Journal of Biological Chemistry. 03 November 1995, Vol. 270, No. 3, pages 26092-26098, see entire document.	1-12
Y	AHMED, M. Two Highly Similar Multidrug Transporters of *Bacillus subtillis* Whose Expression is Differentially Regulated. Journal of Bacteriology. July 1995, Vol. 177, No. 14, pages 3904-3910, see entire document.	1-12